DogoNogo: Creation and Validation of a New Reaction Time Test

Candidate number: 230972

Wordcount: 5490

Date of submission: 21st May 2024

Supervisor: Dominique Makowski

School of Psychology, University of Sussex

Acknowledgements

This project would not have been possible without my supervisor's presence over the last year, from a junior research assistant role to finishing my undergraduate degree, he has made working in science and research not an abstract thought but a tangible reality. It has been a privilege to grow as an aspiring academic whilst seeing him continue to do at his position. So before all, thank you, Dominque Makowski, for providing tutoring and a foundation to work from over the last year.

I would also like to thank my family, friends, flatmates, acquaintances, missed encounters and passing strangers that I have met over the last few years of my degree. If it wasn't for every stepping stone, shared conversation and connection that has accompanied me along the way I do not think I would have managed it. Getting to learn at Sussex University has been an absolute dream come true. To be the first in my family to come to University and have all the experiences I've had in my time here, from the many academics that I've pestered for kernels of wisdom, to the community that connects us all as people at an institution of learning, it has been a great experience. Another thanks go towards my lab mates, especially those who have accompanied me from JRA to now, for all the help and comradery, and another to those who took the time to take part in my study and let me get a chance to conduct research.
Finally, I would like to thank those who have me eager for what comes next, as there is always the next adventure.

Abstract

Reaction time (RT) is a fundamental measure in cognitive psychology, reflecting processing speed and central nervous system efficiency. Traditional RT tasks often suffer from participant disengagement and variability due to boredom and lack of standardization. This study aimed to create and validate a gamified RT test, the DogoNogo game, designed to enhance participant engagement and provide reliable RT measurements. Twenty-five participants (age range 18-48) completed two sessions of the DogoNogo game, which involves responding to a visual stimulus (a barking dog) by pressing a key. The game features scoring and immediate feedback to maintain engagement. Test-retest and split-half reliability were assessed, and correlations between RT and various demographic, personality, and mood factors were examined. The DogoNogo game demonstrated high test-retest reliability and split-half reliability. A strong significant quadratic relationship between inter-stimulus-interval (ISI) and RT was found. Significant gender differences in RT were observed, with males responding faster than females. A significant positive correlation was found between conscientiousness and RT. The DogoNogo game is a reliable tool for measuring RT, offering high engagement through gamification. Future research should address limitations such as sample size and diversity, and further explore the impact of gamification elements on RT reliability.

DogoNogo: Creation and Validation of a New Reaction Time Test

Reaction time (RT) refers to the interval between the onset of a stimulus and the initiation of a motor response. It is a fundamental measure in cognitive psychology and neuroscience, often used to gauge processing speed—the rate at which an individual can perceive, interpret, and respond to information(Salthouse, 1996). RT has been a studied phenomenon since the early days of experimental psychology(Jensen, 2006) and has been studied in a variety of manners from complex choice, recognition or, what this paper will focus on, simple reaction time(Kosinski, 2008). Understanding RT is crucial as it reflects the efficiency of the central nervous system and is linked to various cognitive functions, including attention, memory, and executive control when part of battery tests(Anger et al., 1993; Russell & Pope, 1990). Faster RTs are typically associated with higher cognitive functioning and better overall mental health, making it an essential focus for research in both clinical and non-clinical populations (Deary et al., 2001) as well as findings indicating RT's effect on intelligence(IQ) and mortality(Deary & Der, 2005; Der & Deary, 2003) make RT a simple way to test effects of demographic data such as age or sex(Der, 2006).

Recent findings(Irwin W, 2010; Johnson et al., 1985) demonstrating that we may have gotten slower over the last 100 years, have brought RT and its reliability back into the scientific conversation, though used in several battery tests for cognitive control(Letz, 1991; Russell & Pope, 1990) recent literature has highlighted possible shortcomings(Kadlec et al., 2023) brought about by lack of standardization and confounds when studying what seems to be a simple cognitive function. From caffeine intake and individual differences(Hedge et al., 2022; Ratcliff, 2008; Sainz et al., 2020), to the design of the experiment(Woods et al., 2015), there is a need to create a robust way of studying RT in different conditions, whilst ensuring that the test is resilient to new behavioural factors. Given its broad relevance, developing reliable and engaging RT assessments is a priority for researchers and practitioners alike.

Attention, Boredom and its effects

Attention is a critical factor influencing RT. Sustained attention allows individuals to maintain focus on a task over extended periods, minimizing lapses that could delay responses (Hunter & Eastwood, 2018). Conversely, lapses in attention, often due to boredom, can significantly impair RT. Boredom, defined as a state of low arousal and dissatisfaction, can decrease engagement and results in physical activities (Weich et al., 2022) and cognitive control, leading to slower and more variable RTs (Bieleke et al., 2021). Understanding how these factors interplay is vital for designing tasks that maintain participants' engagement and yield accurate measurements of processing speed.

In experimental psychology, boredom may be contributing to an unknown amount of noise in results(Meier et al., 2024), whether it is through impaired attention, making a task require more effort or disengagement from the task. The influence of attention and boredom on RT has been well-documented in the psychological literature(Pieters, 1985). For example, Thackray(Thackray et al., 1977) found that performance on tasks requiring sustained attention, such as simulated radar control, was significantly influenced by subjective feelings of boredom and monotony. Their study highlighted how higher boredom levels correlated with increased RT variability, suggesting that maintaining participant engagement is crucial for accurate RT measurement. When designing an experiment, especially for those most vulnerable to the effects of boredom, it is important to consider the participant's attention and how best to capture and sustain it, to gather results that reflect true behaviour(Shamay-Tsoory & Mendelsohn, 2019) rather than wandering minds.

Gamification and Reaction Time

Gamification—the application of game-design elements in non-game contexts—has emerged as a promising strategy to enhance motivation and engagement in cognitive tasks (Raed S Alsawaier, 2018; Raed S. Alsawaier, 2018; Alsawaier, 2019; Suh et al., 2018). By incorporating features such as scoring, feedback, and competitive elements, gamified tasks can make repetitive or mundane activities more engaging. This approach has been shown to improve participants' focus and reduce the negative impacts of boredom in educational settings (Rivera & Garden, 2021), thereby potentially enhancing the reliability of RT measurements. There is already research which shows how the features found in games, especially highly competitive ones, affect the RT of gamers(Torner et al., 2019).

The Poisson-Erlang model, discussed by Pieters(Pieters, 1985), supports the idea that incorporating engaging elements in tasks can help manage distraction times and improve overall task performance. The model's application in various tasks demonstrated that reducing distraction and increasing engagement could significantly improve RT consistency and reliability , which affirms other research done into the issue of monotony in RT tasks. Findings by Schupp(Schupp & Schlier, 1972) demonstrate the effect of inter-stimulus-interval (ISI) on RT, showing that when monotonous, participants adapt to these temporal structures, where the size and distribution of ISI have a distinct effect on the speed of RT.

Development of a New Reaction Time Task: The DogoNogo Game

In light of these considerations, the present study aimed to create and validate a new RT test incorporating gamified features. The "DogoNogo" game was developed using the Unity game engine, designed to provide an engaging and interactive way to measure RT. The game features a simple premise where participants respond to a visual stimulus (a barking dog) by pressing a key as quickly as possible. The inclusion of scoring and immediate feedback for slow and quick aims to maintain high levels of engagement and motivation throughout the task.

Trying to design a paradigm which can be used for future research, the game's design also allows for easy modifications, such as adjusting the inter-stimulus interval (ISI) and incorporating additional features like NoGo trials or auditory stimuli, by doing so building off of previous ideas for online test creation software(Mathôt et al., 2012; Stoet, 2017) but one that instead has an already existing community for help creating stimulating games. This flexibility is essential for exploring various aspects of processing speed and cognitive function across different populations and settings.

The primary objective of this study was to validate the DogoNogo game as a reliable measure of RT. This involved assessing its test-retest reliability and split-half reliability to ensure consistency and internal coherence of the RT measurements. Additionally, the study aimed to explore the relationship between RT and various participant characteristics, including age, gender, personality traits, and mood states. By doing so, it sought to provide a comprehensive evaluation of the game's utility, validity and identify potential areas for future research.

Method

Participants

Twenty-five participants (*N* = 25) were recruited for the experiment, through a mix of being approached around campus, study swaps and flyers left around campus with QR codes to sign up(see Appendix 1). Ages ranged from 18 to 48(*M* = 24.28, *SD* = 6.45 years). No names or personal information were gathered about any of the participants, their questionnaire and experiment data were recorded anonymously and kept on the University's Box services.

Materials

Questionaire's

A custom questionnaire was created using jsPsych and presented on the experiment's GitHub page[https://github.com/RealityBending/DoggoNogo]. The questionnaire data is linked to a unique identifier number. The first page is an information and consent form for the experiment. Following the consent form, a series of demographic questions are asked: age, gender, ethnicity and education level. Followed by questions to assess potential confounding factors that might influence reaction time (RT): current fatigue level and caffeine consumption within the past two hours. A series of questions are asked about the participant's exercise and online fast-paced stimuli exposure (e.g., TikTok, video games) over the past two weeks. These questions used 5-point Likert scales ranging from "not at all" to "more than 8 hours worth" and used the two-week timescale to follow the same format as the PHQ-4 questionnaire.

Mini-IPIP-6. The Mini-IPIP6(Milojev et al., 2013) is a short form 24-item measure of six personality traits, five being from the highly used and studied “Big Five” model of Personality including the addition of the Honesty-Humility dimension as suggested by the HEXACO personality model(Ashton & Lee, 2009). In the validation paper for this brief personality measure(Sibley et al., 2011), the correlations between scores at Time 1 and Time 2 (one year apart) ranged from .753 to .983 for the six personality dimensions, indicating good stability of the measure over time. The Mini-IPIP6 was used to reduce experiment time and the risk of boredom to the participant.

PHQ-4. The PHQ-4 is an ultra-brief 4-item measure of anxiety and depression symptoms with growing popularity in research and clinical settings due to its brief and tactful measures. In the validation study(Kroenke et al., 2009) the authors reported a Cronbach's alpha of 0.80 for the PHQ-4, indicating good internal consistency and suggesting that the questions effectively measure anxiety and depression presence as well as severity. The PHQ-4 was elected to be used in this experiment to test for correlations between anxiety and depression symptoms with reaction time, whilst remaining brief.

DogoNogo Game

The DogoNogo game was created in the Unity game engine using C#, with simple retro graphics created in Microsoft Paint or using the free graphic creation tools in Unity. The game source code and build are both fully available on the experiments GitHub, along with a demo and fully commented code to aid in future research conducted with the game, for example, it is simple to change the ISI range, amount of trials or add future features such as a NoGo condition or sound. The game has a title screen and instructions for how to play the game, instructing participants to press the space key when prompted by the image of a dog barking for a sausage, and a box to enter the participant number and session number to track results. Each trial consists of a dog sprite in the “rest” state for a constant wait time, followed by a random wait time(ISI), after which the dog “sprite”/image changes from “rest” to “bark”, the aforementioned image of the dog barking for a sausage which prompts the participants to press the space key. On detection of this press, the ISI for the current trial is recorded along with the reaction time(the time between the prompt stimuli appearing and the space key being pressed), also calculated and recorded is the current rolling median time of all reaction time trials. After the key press the game checks to see if the reaction time for this trial is faster or slower than the participant's average, if it is the same or quicker than the average reaction time so far the dog image/”sprite” returns to the “rest” state if it is slower than the average a red image of the dog is displayed for a second before turning back into the rest condition. This red dog image, referred to in the code as “hurt” dogo, is used as a negative reinforcement, for positive reinforcement a scoreboard exists at the top of the screen, which goes up by 1-1000 depending on how fast the reaction per trial, faster the reaction the higher the score. After this, the next trial begins with the rest state. The game build for this experiment had a constant wait time of one second at the start of every trial, and a random wait time/ISI between 0.1 and 3.5 seconds for every trial, with 120 trials for each session of the game. When checking the game data we computed a mean ISI of 1.80 seconds and a median of 1.82 seconds, showing that the game successfully tested a good range of ISI values. See Figure 1 for the game’s title screen, and the dogo in each of the rest, bark and hurt condition.

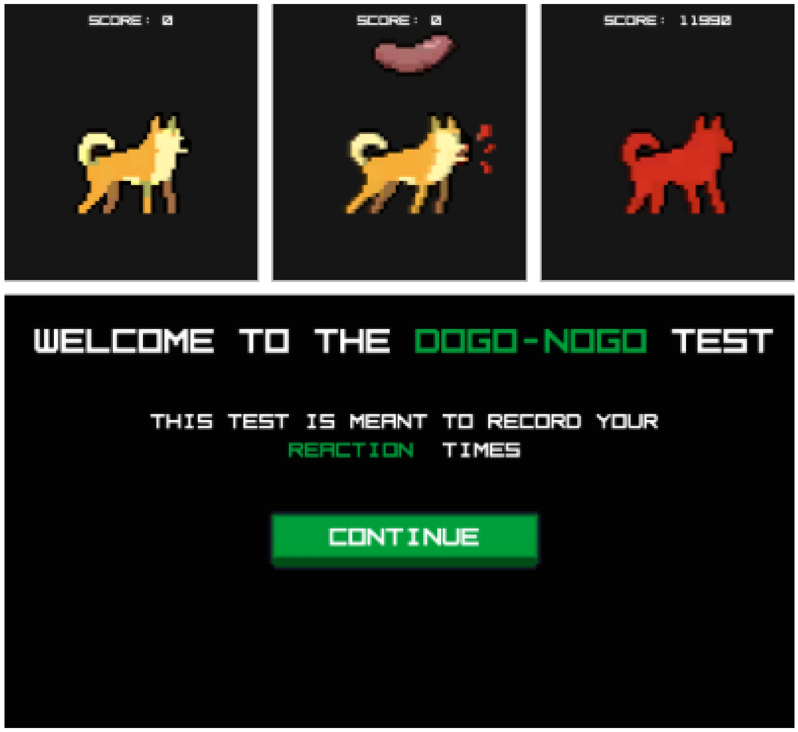


Figure 1: Stimulus "Dogo" sprite in its three conditions, from left to right. Dogo rest, Dogo Bark, Dogo Hurt. Dogo-Nogo intro screen at the bottom

Data Availability

Data was stored anonymously using unique identifiers, no information that may be used to identify the participant was kept. Data was stored on University Box cloud services, with only the experimenter and supervisor of the experiment able to view the data and analysis files.

The game and its source code as well as videos of both are available on the experiments GitHub page[https://github.com/RealityBending/DoggoNogo]. The questionnaire used is also available on the experiment page. Data from the questionnaire is saved as a .json and game data files are saved in the user's AppData folder on the computer used.

Design

This experiment employed a within-subjects design. All participants (n = 25) completed two sessions of a DogoNoGo game, with the independent variable (IV) being the interstimulus interval (ISI) between the "rest dog" and "bark dog" stimuli. The dependent variable (DV) was the participant's reaction time (RT), measured as the time between the appearance of the prompt/"bark dog" stimulus and the participant's space key press. Two sessions were conducted with a break of questionnaires in between to prevent order effects and the participant losing attention and providing inaccurate questionnaire information or slower key presses.

Procedure

A custom questionnaire was created using jsPsych and presented on the experiment's GitHub page. The questionnaire aimed to capture key demographic information and lifestyle habits that may affect or relate to attention and reaction time data. The demographic questions were about age, gender, ethnicity and education level. Lifestyle habits that we tried to capture were sport and exercise, addressed as two separate categories to capture the difference between hobby activities and intense fitness, visual media time was split between time spent playing video games and social media to see if these compromised two distinct types of habits. PHQ and IPIP6 questionnaires were chosen and added to the jsPsych questionnaire due to their short nature and to find mood symptoms and personality data. 5 point Likert scales were used for exercise, virtual stimulation and PHQ, the PHQ was adapted to include a fifth central item for “several days” to see if more nuanced data could be gathered in comparison to the base scale. Participants were able to enter in set numbers for social media and sports to give greater choice to participants. The DogoNogo game was given the parameters as set out in the materials section and built as a .exe game file to run, this was to ensure that for future research it is known what game version and parameters were used for each experiment. The game file and folder were installed on the experiment's computers before the experiment was run.

Participants were either asked to arrive at a certain time or could book in depending on the method of recruitment. After arriving they would be shown to a computer with the questionnaire already loaded from the experiments GitHub page with their assigned participant ID entered in. After clicking continue the participant would be asked to read the information and consent form before deciding to accept and continue the experiment, after which they would work through the demographic and lifestyle habits questions before a screen indicating to let the experimenter know they have completed half the questions. At this point, the DogoNogo game would be loaded, with the experimenter informing the participant they were going to play a game created to test for reaction times, and made aware of the dogo rest state and the stimulus prompt, dogo bark, state to which they should press the space key as quickly as possible. Participants were told that they should try and achieve the highest score they could and if they were going too slow a red dogo image would appear. Participants would then play the game for 120 trials before being shown the end screen, at which point they would be asked to fill in the remaining questions before playing a second session of the game(120 trials again), and being given the exact same instructions apart from a suggestion they should try and beat their score from last time. After the second game session the experimenter would inform the participant the experiment was over and that the red dog was their own average they were competing against, they would also ask the participant if they found the game engaging and how so. The experimenter would then ensure the participant was happy to leave and had no further questions before ending.

Ethical Issues

Ethical approval was obtained from the School of Psychology, University of Sussex (ER/AM2351/3). The study involved no sensitive material or distressing stimuli, the possibility of adding audio to the game was dismissed at this stage due to worries over the effect a distressed dog whine may have had on participants. The questions in the PHQ-4 inquire about the mood state of the participant and depressive symptoms but make no reference to particular distressing events or actions. To ensure no possible ethical concerns, the consent form before the questionnaire outlines all of the material covered in the experiment and the experimenter gives a debrief at the end to check for engagement in the game and for any questions or worries of the participant.

Results

General Relationship between Reaction Time and Inter-State-Interval (ISI)

The relationship between interstimulus interval (ISI) and reaction time (RT) was examined using a quadratic linear model. The model was highly significant, $F(2,5997) = 126.8$, $p < .001$, indicating that ISI significantly predicts RT. The model explained approximately 4.1% of the variance in RT ($R^2 = 0.041$). Examination of the model coefficients showed a highly significant negative linear component ($\beta 1 = -1.514$, $t(5997) = -14.458$, $p < .001$) and a highly significant quadratic component ($\beta 2 = 0.700$, $t(5997) = 6.682$, $p < .001$). These results suggest that the relationship between ISI and RT is not purely linear but exhibits a quadratic nature.

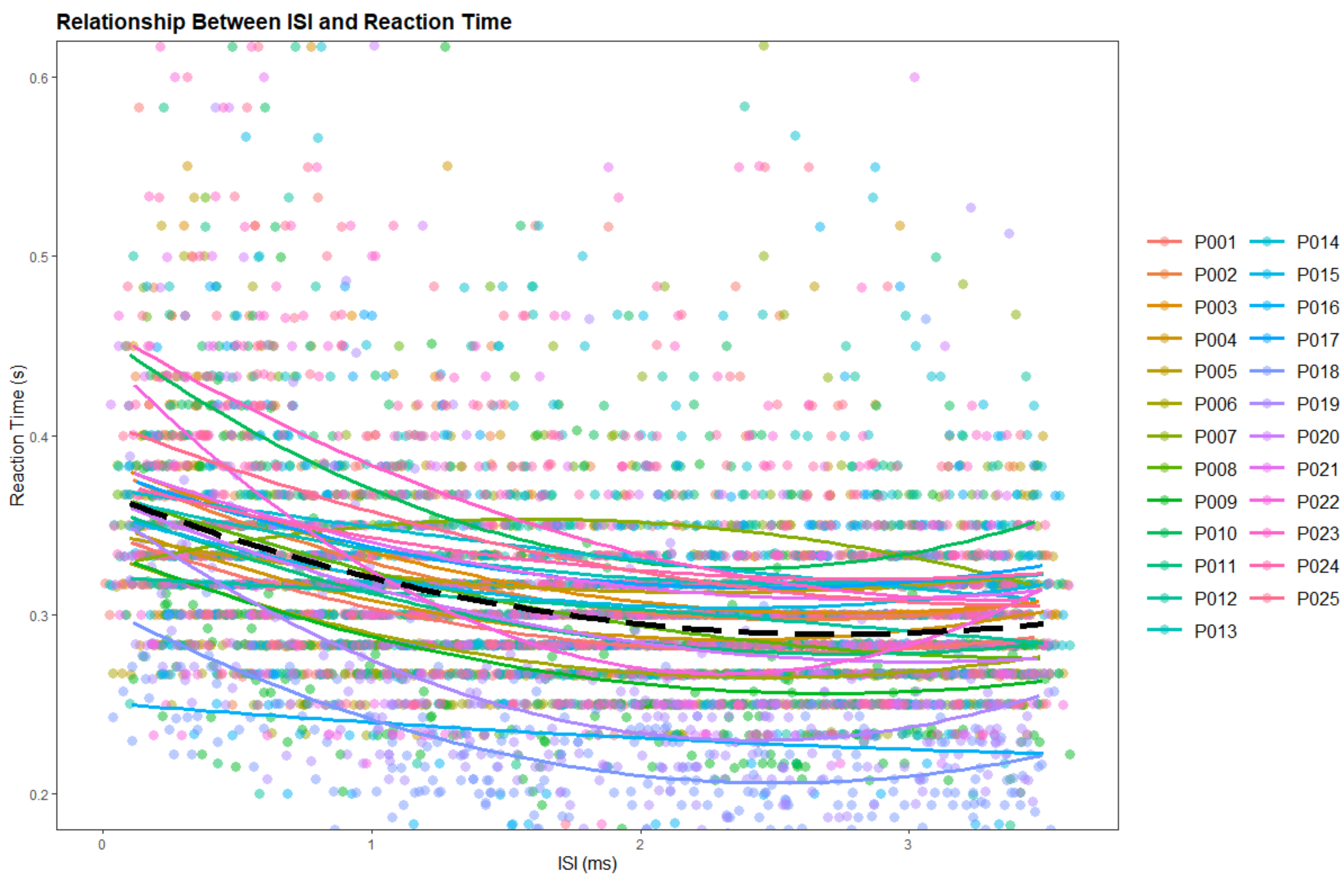


Figure 2: Scatter Plot of Reaction Time against Inters-Stimulus-Interval(ISI). Quadratic regression line's for each participant in colour, overall trend represented by dashed black line

Figure 2 illustrates the relationship between RT and ISI, the overall trend can be seen to reach a low before levelling out and beginning to rise as ISI increases to its max of 3.5 seconds. Individual lines for each participant's regression relationship show that this trend is pronounced in some more than others. The results show that whilst there is a strong significant negative linear relationship between ISI and RT, meaning as ISI increases reaction time decreases, the strong significant quadratic relationship between the two variables demonstrates a more complex relationship than a purely linear one, in which attention may be lost then regained.

Reliability

Two forms of reliability measures were used to ensure the reproducibility of the new reaction time test, test-retest reliability and split-half reliability. To assess the test-retest reliability of the reaction time measurements, Pearson's product-moment correlation was calculated between the mean reaction times in Session 1 and Session 2 for each participant. The results indicated a significant positive correlation, $r(23) = 0.76$, $p < .001$. The 95% confidence interval for the correlation coefficient ranged from 0.52 to 0.89, suggesting a strong and statistically significant relationship between the two sessions. These findings demonstrate that participants' reaction times were consistent over the two sessions, indicating good test-retest reliability.

Split-half reliability was evaluated by randomly dividing the 120 trials of Session 1 into two halves and calculating the mean reaction time for each half. Pearson's product-moment correlation between the two halves was found to be $r(23) = 0.85$, $p < .001$, with a 95% confidence interval ranging from 0.68 to 0.93. This high correlation indicates strong internal consistency within the session, suggesting that the reaction time measurements were reliable across different subsets of trials.

The high test-retest reliability ($r = 0.76$) and split-half reliability ($r = 0.85$) of the reaction time data demonstrate that the measurements are both stable over time and internally consistent. These reliability indices provide strong evidence for the reproducibility and dependability of the reaction time experiment.

Questionaire Correlations

Correlations were computed to explore the relationships between reaction time (RT) and various participant characteristics. Both mean and median RT was used for the analyses to control for irregular data, and the correlations between them were generally high (around 0.8 or above), indicating strong agreement between these measures of central tendency and lack of extreme results for fast or slow reaction times.

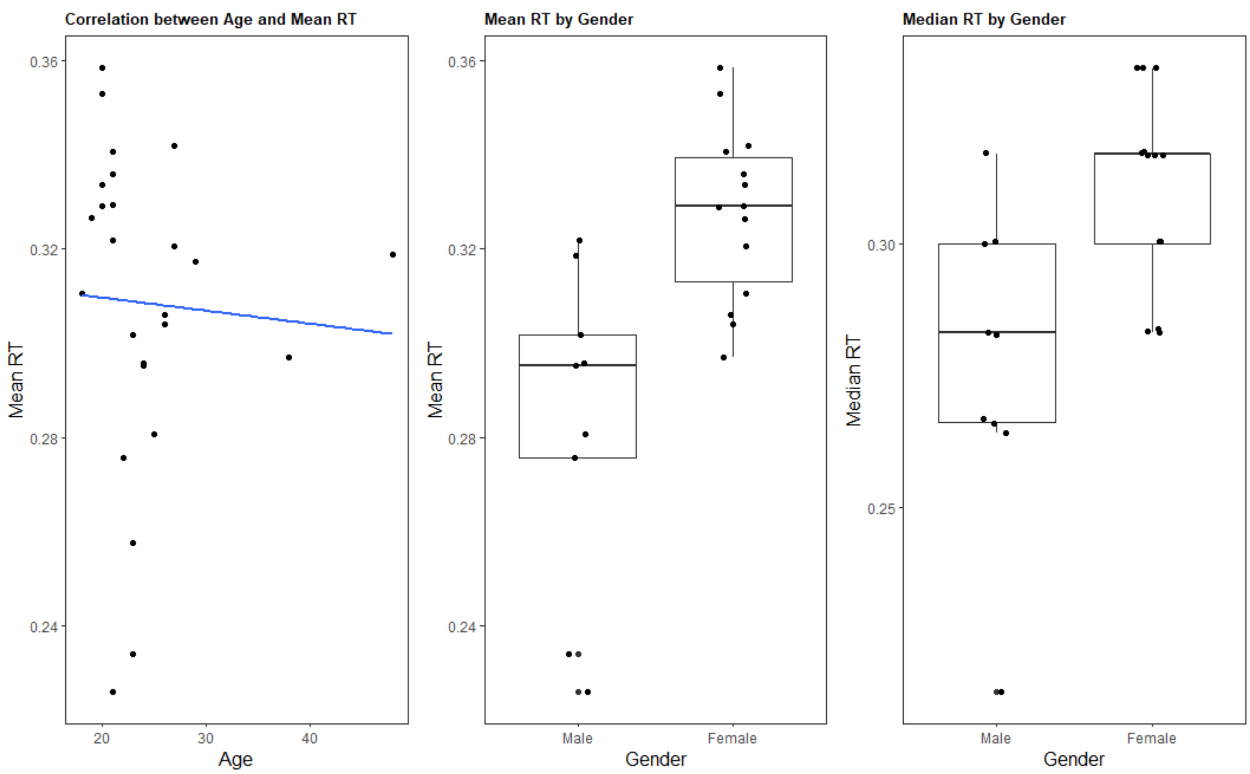


Figure 3: Demographic Information Correlation with Mean and Median RT

Demographics

A Pearson correlation was computed to assess the relationship between age and mean reaction time (Mean RT). There was a very weak, non-significant negative correlation between the two variables, $r = -.05$, $p = .80$, 95% CI [-0.44, 0.35], which was to be expected from a mainly young age sample.

Gender differences in reaction time were assessed using Welch's two-sample t-test. There was a significant difference in Mean RT between males ($M = 0.283$, $SD = 0.022$) and females ($M = 0.328$, $SD = 0.021$), $t(11.07) = -3.60$, $p = 0.004$, 95% CI [-0.071, -0.017]. A similar significant difference was found for median reaction time (Median RT) between males (M = 0.277, SD = 0.021) and females ($M = 0.311$, $SD = 0.018$), $t(11.86) = -3.07$, $p = 0.010$, 95% CI [-0.057, -0.010] (see Figure 3).

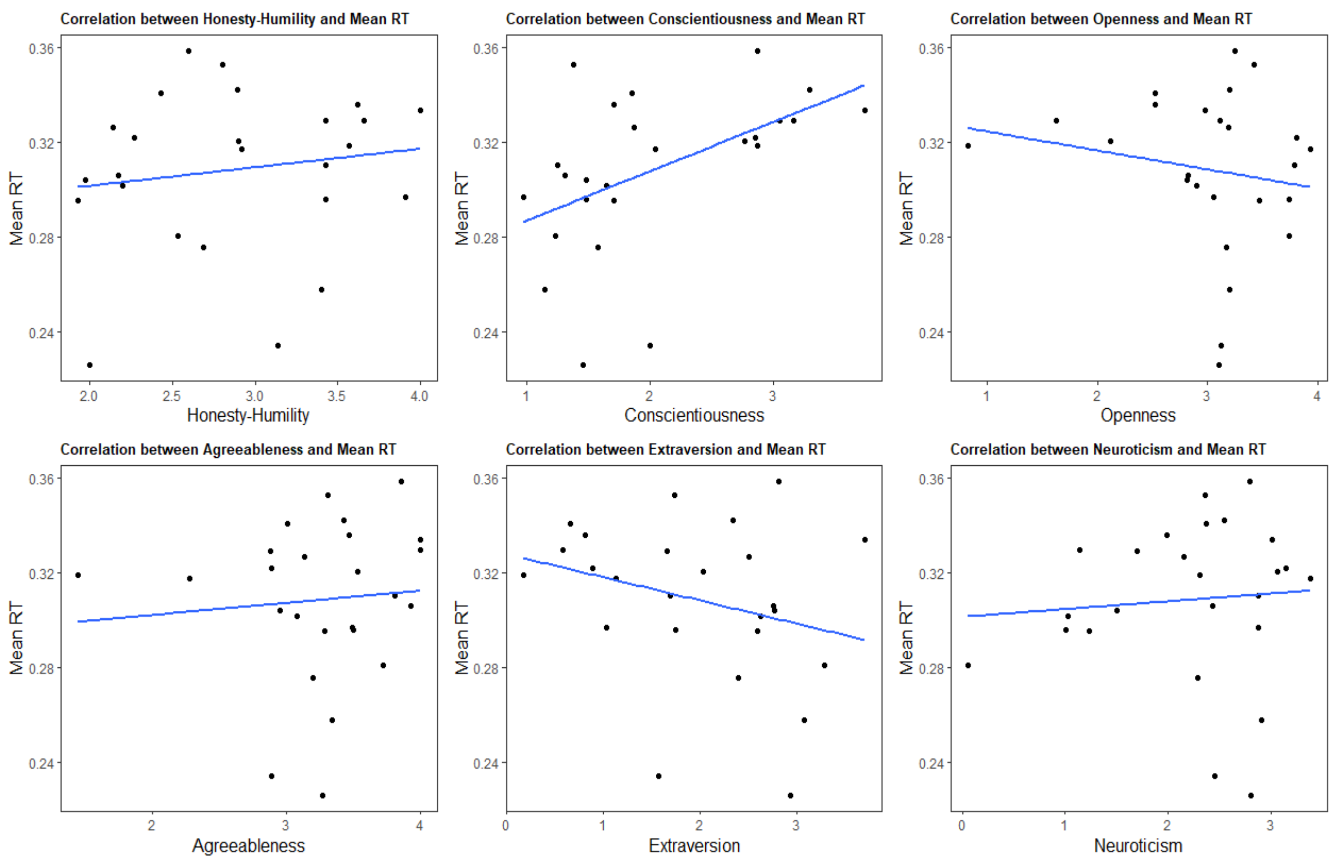


Figure 4: Personality trait scores for the IPIP6 and correlation with Mean RT

Personality

Honesty-Humility (IPIP6)

The correlation between Honesty-Humility and Mean RT was positive but non-significant, $r = 0.15$, $p = .47$, 95% CI [-0.26, 0.51]. For Median RT, the correlation was also positive and non-significant, $r = 0.20$, $p = .34$, 95% CI [-0.21, 0.55] (see Figure 4).

Conscientiousness (IPIP6)

A significant positive correlation was found between Conscientiousness and Mean RT, $r = 0.49$, $p = .012$, 95% CI [0.12, 0.74]. The correlation with Median RT was also highly significant, $r = 0.52$, $p = .008$, 95% CI [0.15, 0.76] (see Figure 4).

Openness (IPIP6)

The correlation between Openness and Mean RT was negative and non-significant, $r = -0.17$, $p = .43$, 95% CI [-0.53, 0.24]. For Median RT, the correlation was also negative and non-significant, $r = -0.12$, $p = .58$, 95% CI [-0.49, 0.29] (see Figure 4).

Agreeableness (IPIP6)

The correlation between Agreeableness and Mean RT was very weakly positive and non-significant, $r = .09$, $p = .68$, 95% CI [-0.32, 0.46]. Similarly, for Median RT, the correlation was also very weakly positive and non-significant, $r = -0.01$, $p = .95$, 95% CI [-0.41, 0.38] (see Figure 4).

Extraversion (IPIP6)

There was a negative, non-significant correlation between Extraversion and Mean RT, $r = -.28$, $p = .18$, 95% CI [-0.61, 0.13]. For Median RT, the correlation was also negative though approached significance, $r = -0.39$, $p = .06$, 95% CI [-0.68, 0.01] (see Figure 4).

Neuroticism (IPIP6)

The correlation between Neuroticism and Mean RT was very weak and non-significant, $r = .08$, $p = .70$, 95% CI [-0.32, 0.46]. For Median RT, the correlation was similarly weak and non-significant, $r = .05$, $p = .82$, 95% CI [-0.35, 0.43] (see Figure 4).

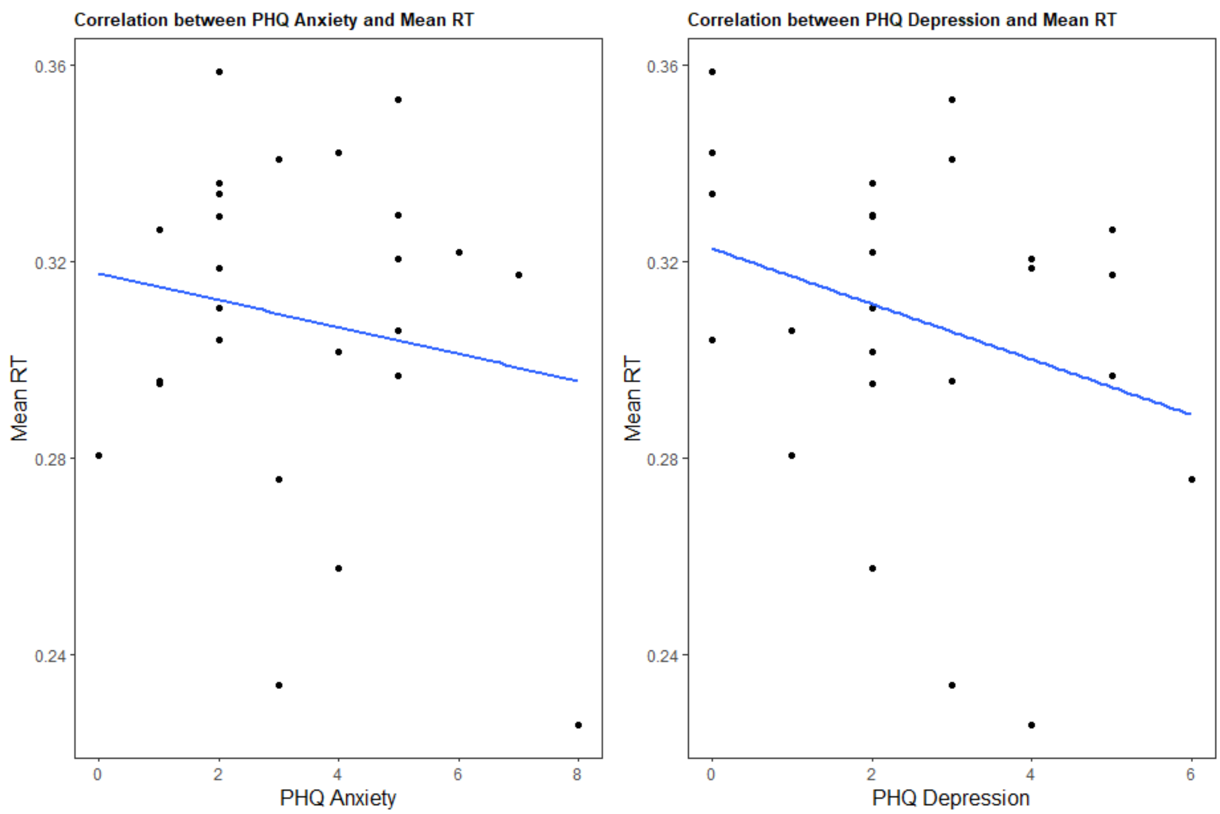


Figure 5: Scores on the PHQ4 test for Anxiety and Depression correlated with Mean RT

Mood

The correlation between PHQ Anxiety scores and Mean RT was negative and non-significant, $r = −0.16$, $p = .43$, 95% CI [-0.53, 0.25]. Similarly, the correlation with Median RT was negative and non-significant, $r = −0.20$, $p = .35$, 95% CI [-0.55, 0.22].

For PHQ Depression scores, there was a non-significant negative correlation with Mean RT, $r = -0.29$, $p = .17$, 95% CI [-0.61, 0.12]. Median RT also showed a non-significant negative correlation, $r = -0.24$, $p = .26$, 95% CI [-0.58, 0.18] (see Figure 5).

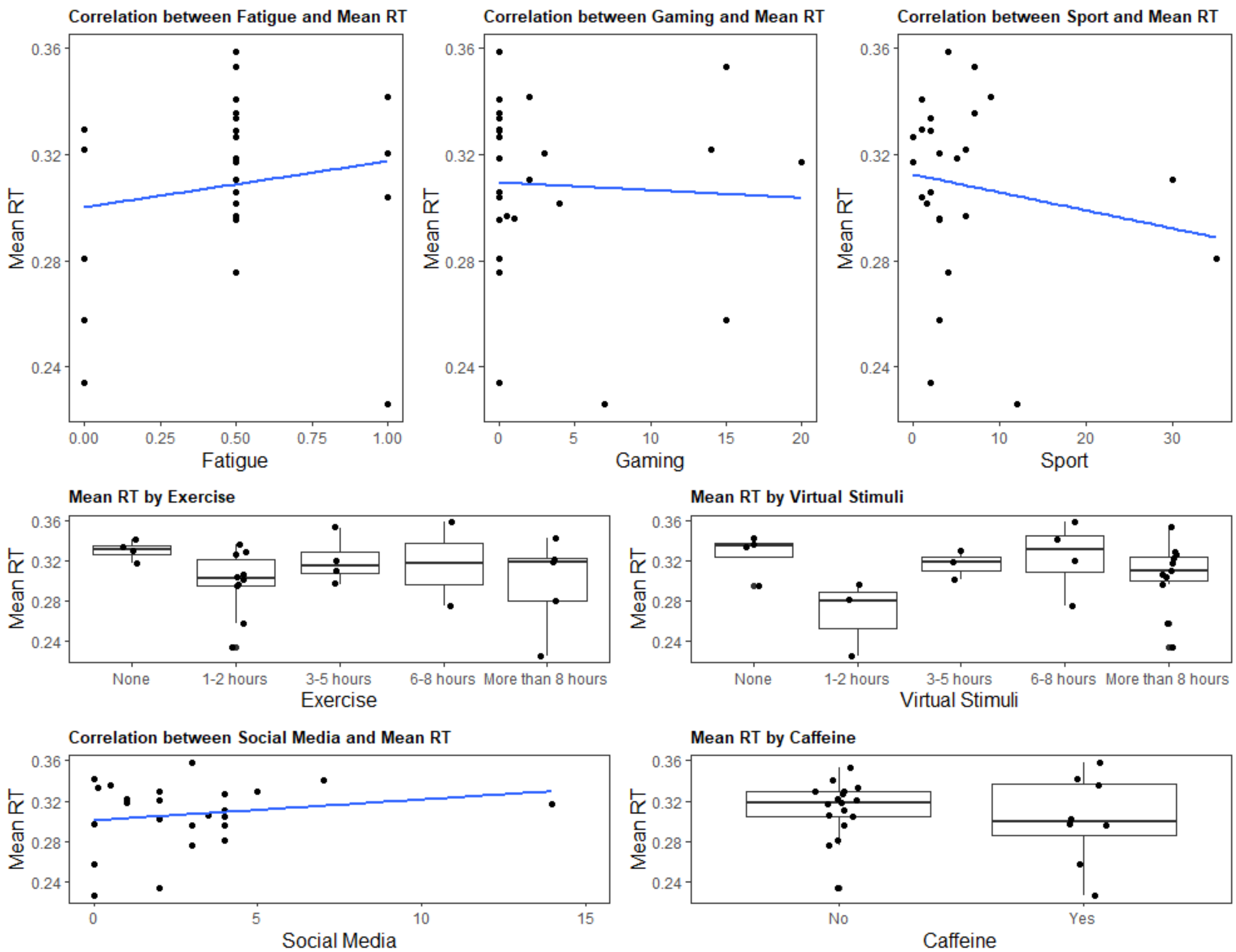


Figure 6: From top left to bottom right; Fatigue level against Mean RT where 0 = not tired 0.50 = somewhat tired 1 = very tired, Hours spent playing video games over last two weeks against mean RT, Hours spent in sports activities over the last two weeks against mean RT, Hours spent exercising over last two weeks against mean RT, Hours spent reacting to virtual stimuli (scrolling on social media, online games) over last two weeks against mean RT, Hours spent on social media over last two weeks against mean RT with scale shortened to 15 hours, Consumption of caffeine within last two hours before experiment against mean RT

Characteristics

Fatigue and Caffeine

A non-significant positive correlation was observed between Fatigue and Mean RT, r = .16, p = .45, 95% CI [-0.25, 0.52], and again for Median RT, r = .04, p = .84, 95% CI [-0.36, 0.43].

The relationship between caffeine consumption and reaction time was examined using Welch's two-sample t-test. There was no significant difference in Mean RT between participants who consumed caffeine (M = 0.302, SD = 0.024) and those who did not (M = 0.312, SD = 0.027), $t(9.77)$ = 0.57, p = .58, 95% CI [-0.028, 0.048]. Median RT also showed no significant difference between caffeine consumers (M = 0.288, SD = 0.021) and non-consumers (M = 0.300, SD = 0.027), $t(8.78)$ = 0.73, p = .48, 95% CI [-0.024, 0.047] (see Figure 6).

Sport and Exercise

For sport participation, the correlation with Mean RT was weakly negative and non-significant, r = −0.17, p = .41, 95% CI [-0.53, 0.24], and with Median RT, r = −0.29, p = .15, 95% CI [-0.62, 0.11].

Exercise frequency showed a weak negative correlation with Mean RT, r = −0.15, p = .49, 95% CI [-0.51, 0.26], and a slightly stronger negative correlation with Median RT, r = −0.26, p = .21, 95% CI [-0.59, 0.15] (see Figure 6).

Gaming and Virtual Stimuli

The correlation between gaming hours and Mean RT was very weakly negative and non-significant, *r* = −0.05, *p* = .81, 95% CI [-0.44, 0.35]. A similar result was observed for Median RT, *r* = −0.03, *p* = .87, 95% CI [-0.42, 0.37].

For virtual stimuli exposure, both Mean RT (*r* = −0.03, *p* = .89, 95% CI [-0.42, 0.37]) and Median RT (*r* = .05, *p* = .80, 95% CI [-0.35, 0.44]) showed non-significant correlations (see Figure 6).

Social Media

A Spearman correlation was computed for social media usage due to the presence of extreme results in the data. The correlation with Mean RT was weakly positive but non-significant, *rho* = 0.18, *p* = .39, and with Median RT, *rho* = 0.25, *p* = .22 (see Figure 6).

For the full results and p-values of each category see Tables 1, 2, 3 and 4 below.

Table 1:

Table showing results for correlation of age and RT and T-Test results for Male and Female reaction Time results with p-values

Variable	Mean_RT Correlation (r)	Mean_RT p-value	Median_RT Correlation (r)	Median_RT p-value	T-test Difference	T-test p-value
Age	-0.0528	0.8021	-0.075	0.7216	-	-
Gender (Female vs Male)	-	-	-	-	-0.0443	0.0042
Gender (Female vs Male)	-	-	-	-	-0.0336	0.0099

Table 2:

Table showing correlation results for personality trait scores on the IPIP6 questionnaire and reaction time with p-values

Variable	Mean_RT Correlation (r)	Mean_RT p-value	Median_RT Correlation (r)	Median_RT p-value
Honesty-Humility	0.1504	0.4731	0.201	0.3354
Conscientiousness	0.4929	0.0123	0.5177	0.008
Openness	-0.1666	0.4261	-0.1177	0.5751
Agreeableness	0.0854	0.6847	-0.0145	0.9452
Extraversion	-0.2772	0.1798	-0.3877	0.0555
Neuroticism	0.0807	0.7014	0.0476	0.8211

Table 3:

Table showing results for correlations between calculated scores for Anxiety and Depression scales of the PHQ4 and reaction time with p-values

Variable	Mean_RT Correlation (r)	Mean_RT p-value	Median_RT Correlation (r)	Median_RT p-value
PHQ_Anxiety	-0.164	0.4333	-0.1965	0.3466
PHQ_Depression	-0.2863	0.1653	-0.2351	0.258

Table 4:

Table showing results for reaction time correlations with Fatigue before the experiment, hours spent gaming, on social media, hours playing sports, hours exercising, hours reacting to virtual stimuli and t-test result for reaction time and if caffeine was consumed within two hours before the experiment. p-values included

Variable	Mean_RT Correlation (r)	Mean_RT p-value	Median_RT Correlation (r)	Median_RT p-value	T-test Difference	T-test p-value
Fatigue	0.1572	0.4531	0.0424	0.8405	-	-
Gaming	-0.0512	0.8078	-0.0333	0.8745	-	-
Sport	-0.1719	0.4113	-0.2945	0.153	-	-
Exercise	-0.1464	0.485	-0.2606	0.2083	-	-
VirtualStimuli	-0.0298	0.8874	0.0543	0.7967	-	-
SocialMedia (Spearman)	0.1785	0.3933	0.2546	0.2193	-	-
Caffeine (Yes vs No)	-	-	-	-	0.0098	0.5805
Caffeine (Yes vs No)	-	-	-	-	0.0116	0.4832

Discussion

The primary objective of this study was to create and validate the DogoNogo game as a reliable measure of reaction time (RT). Our results indicated a significant quadratic relationship between inter-stimulus-interval (ISI) and RT, with a strong linear and quadratic component, suggesting a complex relationship where RT decreases with increasing ISI up to a certain point before levelling out and increasing again. Additionally, the high test-retest reliability (r = 0.76) and split-half reliability (r = 0.85) demonstrated that the DogoNogo game provides stable and consistent RT measurements. However, the analysis showed that only a small proportion of RT variance, around 4.1%, could be attributed to ISI variation. We observed correlations of participants' mean and median RT with demographic, personality, mood and characteristics such as hours spent engaged in sports or social media. Gender differences were observed, with males having significantly faster mean and median RTs compared to females. Correlations with questionnaire data revealed a significant positive relationship between conscientiousness and RT, and a negative relationship with extraversion that approached significance, whilst other personality traits and mood states showed non-significant correlations with RT though positive and negative trends were observed.

This study contributes to the existing literature by validating a gamified RT task that incorporates features to maintain engagement and minimize the effects of boredom. Our findings on the relationship between ISI and RT align with previous research, but the observed quadratic relationship adds a new dimension to our understanding of this relationship. The significant gender differences in RT observed in our study are consistent with prior research, further validating the robustness of our findings. The practical implications of our findings are substantial. This is particularly useful in contexts where maintaining attention is crucial for obtaining accurate measurements in research, educational or clinical populations.

The observed gender differences in RT also suggest that gender-specific norms may be necessary for interpreting RT data accurately. However, previously found results that indicate caffeine and exercise may impact RT were not observed here, and it should be noted that to better understand individual differences a higher sample size should be studied to properly asses the relationships of RT with lifestyle choices. For demographics too, we dropped those who identified outside the gender binary due to low sample size as well as age analysis due to the majority university student age sample, though having greater variation of gender and age may provide insights into their effect on individuals' processing speed. No significant correlations were found for either of the mood factors from the PHQ test, however, negative trends were seen in both, if future research were to be conducted with this paradigm it would be an advantage to see if this trend is found again to find potential links between clinical symptoms and simple RT measures. So too for personality correlates, whilst most personality traits scores uncovered by the IPIP6 provided non-significant trends with RT, conscientiousness showed a highly significant positive correlation with RT, possibly showing the more someone exhibits conscientious traits, to be careful or diligent, the slower one's reaction time is. Similarly, for extraversion, a negative relationship with RT was observed that approached significance, showing that perhaps the more outgoing and assertive traits one has, the quicker one may be with processing and reacting to information. Additionally, potential confounding factors, such as individual differences in familiarity with video games, were not fully controlled for and may have influenced the results.

Future research should aim to address these limitations by increasing the sample size and ensuring a more diverse participant pool. Further studies could explore the effects of varying ISI ranges more extensively, by increasing the max ISI to observe lapses in attention, and test the DogoNogo game in different populations, such as older adults or clinical groups. Additionally, examining the impact of different gamification elements, such as varying types of feedback or competitive elements, could provide deeper insights into how to enhance engagement and reliability in RT tasks. We found that after both sessions of the game when asked, participants would note how invested they were to get a high score and avoid getting too many red dogs, and after being told that they were trying to “beat” their own average score many laughed at their own frustrated engagement. It would be more beneficial to operationalise this “gamification” measure to better study the effectiveness of game elements in psychological measures of attention and engagement.

In summary, this study successfully validated the DogoNogo game as a reliable measure of RT, demonstrating high test-retest and split-half reliability. The findings underscore the complexity of the relationship between ISI and RT and highlight the importance of gamification in maintaining participant engagement and obtaining consistent measurements. The open-source nature of the DogoNogo game and its development in the Unity engine make it easily modifiable for future research, ensuring the standardization of test apparatus and facilitating the creation of new experiments. This flexibility, combined with the demonstrated reliability of the game, positions DogoNogo as a valuable tool for advancing cognitive research in diverse populations and settings.

References

Alsawaier, R. S. (2018). The effect of gamification on motivation and engagement. *The International Journal of Information and Learning Technology*, *35*(1), 56-79.

Alsawaier, R. S. (2018). *The Effect of Gamification on Motivation and Engagement in Three WSU College Courses*

Alsawaier, R. S. (2019). Research trends in the study of gamification. *The International Journal of Information and Learning Technology*, *36*(5), 373-380. https://doi.org/10.1108/IJILT-12-2017-0119

Anger, W. K., Cassitto, M. G., Liang, Y. X., Amador, R., Hooisma, J., Chrislip, D. W., Mergler, D., Keifer, M., Hörtnagl, J., Fournier, L., & et al. (1993). Comparison of performance from three continents on the WHO-Recommended Neurobehavioral Core Test Battery. *Environ Res*, *62*(1), 125-147. https://doi.org/10.1006/enrs.1993.1097

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, *91*(4), 340-345.

Bieleke, M., Barton, L., & Wolff, W. (2021). Trajectories of boredom in self-control demanding tasks. *Cognition and Emotion*, *35*(5), 1-11. https://doi.org/10.1080/02699931.2021.1901656

Deary, I. J., & Der, G. (2005). Reaction Time Explains IQ's Association with Death. *Psychological Science*, *16*(1), 64-69. https://doi.org/10.1111/j.0956-7976.2005.00781.x

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, *29*(5), 389-399. https://doi.org/https://doi.org/10.1016/S0160-2896(01)00062-9

Der, G. (2006). Reaction time age changes and sex differences in adulthood. Results from a large, population based study: the UK Health and Lifestyle Survey. *Psychol Aging*, *21*, 62-73.

Der, G., & Deary, I. J. (2003). IQ, reaction time and the differentiation hypothesis. *Intelligence*, *31*(5), 491-503. https://doi.org/https://doi.org/10.1016/S0160-2896(02)00189-7

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2022). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(10), 1448-1469. https://doi.org/10.1037/xlm0001028

Hunter, A., & Eastwood, J. D. (2018). Does state boredom cause failures of attention? Examining the relations between trait boredom, state boredom, and sustained attention. *Experimental Brain Research*, *236*(9), 2483-2492. https://doi.org/10.1007/s00221-016-4749-7

Irwin W, S. (2010). Simple reaction time: It is not what it used to be. *The American Journal of Psychology*, *123*(1), 39-50. https://doi.org/10.5406/amerjpsyc.123.1.0039

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier.

Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, *40*(8), 875.

Kadlec, J., Walsh, C., Sade, U., Amir, A., Rissman, J., & Ramot, M. (2023). *Putting cognitive tasks on trial: A measure of reliability convergence*. https://doi.org/10.1101/2023.07.03.547563

Kosinski, R. J. (2008). A literature review on reaction time. *Clemson University*, *10*(1), 337-344.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ–4. *Psychosomatics*, *50*(6), 613-621. https://www.sciencedirect.com/science/article/pii/S0033318209708643?via%3Dihub

Letz, R. (1991). Use of computerized test batteries for quantifying neurobehavioral outcomes. *Environ Health Perspect*, *90*, 195-198. https://doi.org/10.1289/ehp.90-1519520

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Meier, M., Martarelli, C. S., & Wolff, W. (2024). Is boredom a source of noise and/or a confound in behavioral science research? *Humanities and Social Sciences Communications*, *11*(1), 368. https://doi.org/10.1057/s41599-024-02851-7

Milojev, P., Osborne, D., Greaves, L. M., Barlow, F. K., & Sibley, C. G. (2013). The Mini-IPIP6: Tiny yet highly stable markers of Big Six personality. *Journal of Research in Personality*, *47*(6), 936-944.

Pieters, J. P. M. (1985). Reaction time analysis of simple mental tasks: A general approach. *Acta Psychologica*, *59*(3), 227-269. https://doi.org/https://doi.org/10.1016/0001-6918(85)90046-0

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging*, *23*(4), 900-916. https://doi.org/10.1037/a0013930

Rivera, E. S., & Garden, C. L. P. (2021). Gamification for student engagement: a framework. *Journal of Further and Higher Education*, *45*(7), 999-1012. https://doi.org/10.1080/0309877x.2021.1875201

Russell, P. E. F. R. W., & Pope, A. M. (1990). *Behavioral measures of neurotoxicity*. National Academies Press.

Sainz, I., Collado-Mateo, D., & Coso, J. D. (2020). Effect of acute caffeine intake on hit accuracy and reaction time in professional e-sports players. *Physiology & Behavior*, *224*, 113031. https://doi.org/https://doi.org/10.1016/j.physbeh.2020.113031

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, *103*(3), 403.

Schupp, W., & Schlier, C. (1972). The dependence of simple reaction time on temporal patterns of stimuli. *Kybernetik*, *11*(2), 105-111.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, *14*(5), 841-859. https://doi.org/10.1177/1745691619856350

Sibley, C. G., Luyten, N., Purnomo, M., Mobberley, A., Wootton, L. W., Hammond, M. D., Sengupta, N., Perry, R., West-Newman, T., & Wilson, M. S. (2011). The Mini-IPIP6: Validation and extension of a short measure of the Big-Six factors of personality in New Zealand. *New Zealand Journal of Psychology*, *40*(3).

Stoet, G. (2017). PsyToolkit:A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24-31. https://doi.org/10.1177/0098628316677643

Suh, A., Wagner, C., & Liu, L. (2018). Enhancing User Engagement through Gamification. *Journal of Computer Information Systems*, *58*(3), 204-213. https://doi.org/10.1080/08874417.2016.1229143

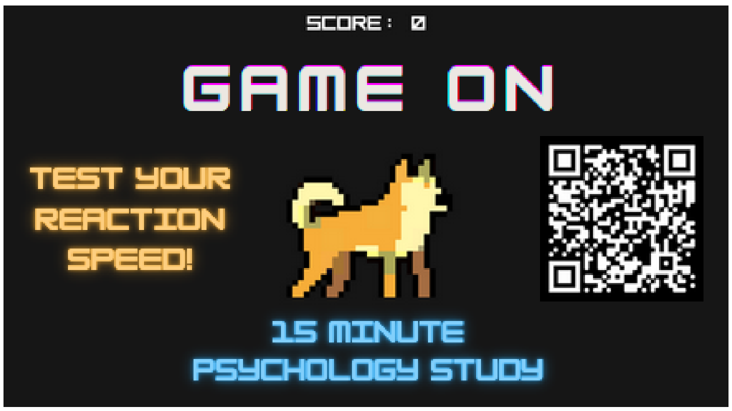
Thackray, R. I., Bailey, J. P., & Touchstone, R. M. (1977). Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task. In *Vigilance: Theory, operational performance, and physiological correlates* (pp. 203-215). Springer.

Torner, H., Carbonell, X., & Castejón, M. (2019). A comparative analysis of the processing speed between video game players and non-players. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport*, *37*, 13-20. https://doi.org/10.51698/aloma.2019.37.1.13-20

Weich, C., Schüler, J., & Wolff, W. (2022). 24 Hours on the Run—Does Boredom Matter for Ultra-Endurance Athletes' Crises? *International Journal of Environmental Research and Public Health*, *19*(11), 6859. https://doi.org/10.3390/ijerph19116859

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time [Original Research]. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00131

Appendecies



Appendix 1: Recruitment poster