KPU Registry ID Number 1056 **Date Submitted: 29th April 2020**

Study Registration for the KPU Study Registry

Three confirmatory analyses of precognition and micro-pk data gathered using online methods

4/29/20

Authors: Julia Mossbridge & Dean Radin
Submitter: Julia Mossbridge
Contact: jmossbridge@gmail.com

Background

In an exploratory study (some results presented at the 2019 Society for Scientific Exploration meeting), we examined data from the Psi3 smartphone application (the app was originally launched as "PsiQ") as well as the website www.thepremonitioncode.com. All datasets were closed as of April 30, 2019. In addition to smaller effects, we found interesting results based on the analyses of data from one of the smartphone app games, providing apparent evidence for micro-psychokinesis, and we also found interesting support for precognition from the website data. Soon it will be a year from that original cutoff date, and because more participants are using both tools, we thought it useful to perform confirmatory analyses on the new data obtained from May 1 2019 to April 30 2020. This registry document briefly describes the confirmatory hypotheses and their predicted outcomes as well as the planned analyses; detailed methods and analysis information can be obtained once we publish the journal article based on these data. No analyses in the new data set (ending April 30 2020) have yet been performed.

Confirmatory hypotheses & predicted outcomes

<u>Micro-pk/Heart Quest (Psi3)</u>

Original results from exploratory experiment. Each trial of this game produces 2 random reference bits (at the beginning of each game) and 2 random trial bits (each time the user touches the smartphone screen). The goal is to use micro-pk/mental intention to make a robot's heart glow. This is done by matching the trial and reference bits; trial and reference bits are created by a true RNG using the same function in the software that drew on a combination of the accelerometer in the user's phone and a built-in pseudorandom number generator. We found significant psi-missing in this task. The most impressive results (small but significant) from the exploratory analysis were that both of the trial bits were equally likely to be a 0 versus 1 while both of the reference bits were more likely to be a 0 versus 1, an effect especially obvious in the second reference bit (nondirectional Cohen's *g*: first bit: 0.002, second bit: 0.005; note that there were over 300,000 trials so for example in the case of the second bit, $p<0.0\ e^{-10}$). Additionally, we found a gender difference here – women showed a large and significant tendency toward more zeros than ones in both reference bits (Cohen's *g*s: 0.005, 0.011) while men only showed a small effect in the first reference bit and no effect in the second reference bit (Cohen's *g*s: 0.003, 0.001). Further, the proportion of zeros in the second reference bit was significantly greater for women as compared to men (Chi-square=11.547, df=1).

Confirmatory hypothesis. This task captures a micro-pk gender difference between self-reported women and men.

Predicted outcome. All participants will again show a significantly higher number of zeros in the first and second reference bits than would be expected by chance. Self-reported women will again produce trials in which there are significantly more zeros than ones in both the first and second bits than expected by chance, while self-reported men will again show lesser effects as compared to women. There will again be a significant gender difference in the proportion of zeros in the second reference bit, favoring women (see planned analyses).

<u>Precognition/Positive Precog Training/Testing (thepremonitioncode.com)</u>
Original results from exploratory experiment. The software asks the users to choose between two graphs showing the presence or absence of 8 potential elements in an upcoming image. After the user selects one of the graphs, the software uses a true random number generator based on network traffic to choose one of the two images represented by the graphs to display to the user. This is the target image that the user would be aiming to predict if they hoped to get the trial "correct." Results were unimpressive overall (slightly above chance). There was a large primacy bias toward choosing the graph presented on the left side or the top of the screen (depending on the device/browser used). Analyzing only choices made for the right or bottom image, we selected the top 8 images most often incorrect and the top 8 images most often correct. Supporting the idea that certain targets are more "psi-attractive," judges rated the most-likely-correct images as significantly more interesting, on average, than the most-likely-incorrect images (Cohen's *d*: 0.255). Taking a different statistical approach and comparing judge ratings in each questionnaire (without collapsing across questionnaires; see planned analyses) reveals a more robust effect (Cohen's *d*: 3.99).
Confirmatory hypothesis. Images that are subjectively more interesting draw precognitive attention and uninteresting images do not draw precognitive attention, resulting in more hits with interesting targets than uninteresting targets.
Predicted outcome. As we found in the exploratory analysis, among targets presented in the right-hand or bottom position, the top 8 most-often correct targets will be rated as more interesting than the top 8 least-often correct targets (see planned analyses).

Planned Analyses
<u>Overall</u>
We will include in these analyses all complete trials performed during the time period (midnight GMT May 1, 2019 to midnight GMT May 1, 2020) regardless of how many trials were performed by any particular user. Data will be analyzed using Microsoft Excel/Office Libre for Mac, R for Mac, and Matlab 2018b. Alphas will be set at 0.05. Two-tailed null hypothesis significance testing will be used. We will use the same data cleaning and statistical analysis methods we used when we analyzed the initial data sets in the exploratory experiment; these are briefly described for each of the two types of data sets below.
<u>Micro-pk/Heart Quest (Psi3)</u>
Data cleaning. We will remove all incomplete trials (where a response was not made). Because we cannot analyze data from children since our consent form was approved for participants 18 and over, and because we assume participants stating that they are over 100 are not being truthful, data from participants who do not list their age as between 18 and 100 years old (inclusive) will also be excluded. When examining the influence of gender, we will ignore trials performed by participants who did not rate themselves either below 0.25 (male) or above 0.75 (female) on the gender slider; we will then consider self-reported men as those who rate themselves below 0.25 and women as those who rate themselves above 0.75. We will not single out all the trials performed by each participant but instead treat all trials as single data points.

Statistical Analyses

With regard to the number of zeros in the first and second reference bit, the predicted outcome is that all trials when taken together, without respect to gender, will reveal a significantly higher number of zeros in the first and second reference bits than would be expected by chance. Note that this is the direction of the expected effect, though with our test we could also detect significantly fewer zeros than expected by chance. That's because we will use two-tailed (e.g., two-direction) binomial tests to examine potential deviations from randomness for each of the two reference bits. The original effect will be considered to be weakly replicated if only the first or second reference bit reveals a significant effect in this direction, and strongly replicated if both do.

With regard to the predicted gender difference, the predicted outcome is that trials from women will show significantly more zeros than chance in these two reference bits, and that this effect will be significantly bigger in women than men in the second reference bit. For trials from participants of each self-reported gender, we will use binomial tests to examine whether the number of zeros in each reference bit deviates from chance. We predict that the effect sizes (Cohen's *g*) for each reference bit will be in the predicted direction (more zeros) and also will be greater in women than men and that there will be more significant comparisons versus chance (*p*-values from binomial tests) in women versus men. If this occurs, we will consider that the original gender difference is weakly replicated. However, will also use a Chi-squared test to determine whether the number of zeros in the second reference bit differs between genders. If this test is significant and in the direction predicted, and if the weak replication also occurred, we will consider this gender interaction a strong replication of the gender difference in what appears to be a micro-pk effect on the second reference bit.

Assuming a Cohen's *g* of 0.005 (true proportion of zeros – null hypothesis proportion [0.5]=0.005), 80% power and an alpha of 0.05, the number of trials needed to show significance for the binomial test analyses should be > 79,000. If we do not get at least this number of trials, we will consider quantitative trends with the awareness that we are not able to fully test the over-arching hypothesis. The largest effect is the significantly greater number of zeros versus chance obtained in trials from self-reported women. Assuming a Cohen's *g* of 0.011, 80% power and an alpha of 0.05, the number of trials needed to show significance for the binomial test analyses should be > 16,000. If we don't obtain at least this many trials performed by women, we will also consider quantitative trends with the awareness that we are not able to fully test the over-arching hypothesis, and we will do the same with the Chi-squared comparison with men.

<u>Precognition/Positive Precog Training/Testing (thepremonitioncode.com)</u>

Data cleaning

We will include all trials from participants who signed the electronic informed consent form and completed at least one practice or testing trial on the site during the time period, who reported that they were 18 or over, 100 and under, and who were not part of the site development, design, and testing team.

Statistical Analyses

We will assign a score to each trial (0=incorrect, 1=correct), then examine only trials in which the correct targets were presented in the second (right-hand or bottom) position (as in the previous analysis we expect a primacy bias and want to reduce noise in the data set). We will identify the n targets which, when presented in the second position, have a ratio of incorrect to correct trials of 0.4 or less – these are the *top-n-likely-correct targets*. Targets presented in second place were more often incorrect than correct (probably due to the first-place bias), so we will find targets with an incorrect-to-correct ratio of 2.2 or greater – these are the *top-n-likely-incorrect targets*. There will always be the same number of likely-correct- and likely-incorrect-targets in our final data set, and we will exclude targets in one group or the other based on a ranking of their incorrect-to-correct ratios. For example, if we can only find 6 likely-correct targets, will will only use the top 6 likely-incorrect targets in our dataset (to facilitate pairing, see below).

We will then create n unique questionnaires designed to assess the "interestingness" of the targets (n=number of likely-correct and likely-incorrect targets). Across all n questionnaires, each of the top-n-likely-correct targets will be presented with each of the top-n-likely-incorrect targets in different pair combinations across the questionnaires (e.g., the first most likely-incorrect target would be presented with the first most likely-correct target in one questionnaire and the same target would be presented with the second most likely-correct target in another questionnaire, and so on). The n questionnaires will be presented one at a time to different groups of paid workers on Amazon Mechanical Turk, each group consisting of 10 paid workers. Their task will be to look at each of the pairings of the images in the questionnaires and ask themselves which of the two images is most interesting/intriguing, or whether they couldn't determine the answer. Workers will not be told about the context of the original experiment or that it had anything to do with precognition – they will just be asked to compare the images on the basis of interestingness. Each target will be given n independent "interestingness" scores based on the data from each of the n questionnaires. Interestingness scores will be calculated according to the proportion of workers completing a given questionnaire who choose a target as most interesting (1=all workers think it was the most interesting target; 0=none of them think it was most interesting; 0.5=half think it is most interesting).

We then will use between-groups t-tests to compare interestingness scores between the top-n-likely-correct and likely-incorrect in two ways: 1) interestingness scores averaged across all instances of an image from all questionnaires (n values in each group, averaged across all participants and all questionnaire contexts), and 2) interestingness scores averaged across participants within a single questionnaire context (n*n values in each group – note that these are independent observations because each observation is made in the context of comparison with a different image). Note that it is the case that in a single questionnaire (single judge) with a single pairing of images (one question on the questionnaire) if one image is rated interesting, the other is rated uninteresting – this is a dichotomous variable. But averaging the scores for that single image across multiple participants and multiple questionnaires (analysis 1) or across multiple participants within a single questionnaire (analysis 2) both provide continuous variable sets for which between-group tests become appropriate. A significant t-test resulting from either the first or second method of analysis will be considered a replication of the original effect if it is in the direction of greater interestingness scores in the top-n-likely-correct group; if both are significant this will be considered a stronger replication.

Note that if we cannot find at least 7 likely-correct- and likely-incorrect-targets, we will forfeit the first analysis (this is the minimum sample size for the more conservative statistical analysis according to a power calculation using Cohen's *d* of 0.255 with 80% power and an alpha of 0.05). However, for the second analysis using the same calculation we would only require a minimum of 15 scores, which indicates that we would need at least 4 likely-correct- and likely-incorrect-targets (n=16) for that test. So we will perform the second analysis only if n>=4.