Brain and Cognition 146 (2020) 105638

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Accuracy and neural correlates of blinded mediumship compared to controls on an image classification task

A. Delorme[a,b,*], C. Cannard[a], D. Radin[a], H. Wahbeh[a,b]

[a] Institute of Noetic Science, Petaluma, CA, USA
[b] Swartz Center for Computational Neuroscience, Institute of Neural Computation, University of California, San Diego, La Jolla, CA, USA

ARTICLE INFO

Keywords:
Intuition
Mediumship
Electroencephalography
Behavior
Machine-learning

ABSTRACT

In this study, a classification task asked participants to look at 180 facial photographs of deceased individuals (photographs were taken years prior to their deaths) and guess the cause of death from three equiprobable categories: heart attack; death by firearm; or car accident. Electroencephalogram (EEG) and electrocardiogram (ECG) data were simultaneously collected during the task. The participants included individuals who claimed "mediumistic" (psychic) abilities and controls who claimed no mediumistic ability. Pooled data showed accurate guesses for the cause of death (partial $\eta^2 = 0.12$; $p = 0.004$), and control subjects were primarily responsible for this effect (partial $\eta^2 = 0.11$; $p = 0.005$). EEG and ECG differences were found between the mediums and controls. Control participants had larger amplitude event-related potentials (ERP) following the presentation of the images than the mediums, between 80 and 110 ms, and between 200 and 350 ms. This could be interpreted as reflecting greater attention and less response inhibition by controls as compared to the mediums. Participants in the control group also had lower average heart rates than the mediums, possibly indicating less stress during the task. Speculations and limits regarding why controls performed better than mediums are discussed.

1. Introduction

Throughout history and across cultures, some individuals have claimed the ability to access accurate information by non-conventional means (beyond physical senses). Individuals who are allegedly able to access accurate information about deceased individuals are often called "Mediums." This phenomenon has become more popular as mediums are sought out by the bereaved for an understandable wish to remain in contact with their deceased loved ones (Beischel, Mosher, & Boccuzzi, 2015). Beyond that, some of the information provided by mediums is verifiably accurate (Beischel & Schwartz, 2007; Beischel & Zingrone, 2015; Delorme et al., 2013; Rock, Beischel, & Cott, 2009). The verification of information received through mediums was observed often enough that even William James, the "Father of American Psychology," found the topic worthy of scientific investigation (Knapp, 2017).

In a standard "reading" with a medium, a client asks questions about a deceased individual (DI) - for example a friend or relative. The medium purports to communicate with the DI and reports any received information to the client. Mediums are often eager to provide verifiable information to demonstrate that they are indeed communicating with the DI, as requested. Double and triple-blinded experimental protocols (Beischel & Schwartz, 2007; Beischel & Zingrone, 2015; Delorme et al., 2013; Rock et al., 2009) have been employed to verify that information delivered by some mediums is indeed accurate under conditions that exclude mundane explanations and have provided the opportunity to measure behavioral and physiological effects associated with the retrieval of accurate information.

In a previous project, we analyzed brain activity, skin conductance, respiration, heart rate, and peripheral blood flow of mediums while they performed readings under double-blind conditions (Delorme et al., 2013). All four mediums in that study produced results above chance. In three out of the four mediums, the results were independently statistically significant at $p < 0.05$, with a 17.5% average difference between target and decoy readings across the four mediums. One medium showed a decrease in EEG frontal midline theta waves during accurate responses, suggesting a possible decrease in executive functions associated with successful responses (Delorme et al., 2013). A limitation of that experiment was its low statistical power due to the relatively small number of participants and number of trials.

A second experiment with increased statistical power was conducted to evaluate if mediums could accurately classify photographs of living versus deceased individuals under blinded conditions (Delorme, Pierce,

* Corresponding author at: Institute of Noetic Science, Petaluma, CA, USA.
E-mail address: arnodelorme@gmail.com (A. Delorme).

https://doi.org/10.1016/j.bandc.2020.105638
Received 25 March 2020; Received in revised form 14 October 2020; Accepted 19 October 2020
Available online 6 November 2020
0278-2626/

Michel, & Radin, 2018). Four hundred and four images of faces were displayed on a computer screen (half were of deceased and half were alive individuals). Twelve mediums were asked to press one of three keys to indicate if they thought the depicted person was: (1) deceased; (2) living; or (3) unsure if they were alive or deceased. Five of the 12 mediums performed above chance levels. Also, early brain activity, 100 ms after the presentation of the image, significantly differed between the correctly vs. incorrectly categorized images of deceased individuals.

The Delorme et al. (2018) experiment had two caveats. First, non-medium controls were not tested, so we could not assess if mediums were more accurate than people who did not claim mediumship skills. Second, information about whether a given individual in the database was alive or deceased could not be absolutely verified for accuracy as we relied on a social media website to collect that information. Therefore, some individuals who were classified as living may have been deceased at the time of the experiment. Also, regular checks, determining if the living individuals in the database were still alive, were not feasible, which precluded other laboratories from using the same database to replicate our results.

The current study aimed to address these limitations in two ways. First, both mediums and controls were included, and second, all of the photographs were of deceased individuals. The task was to indicate the cause of death out of three possible choices: heart attack; death by firearm; or car accident. Three hypotheses were then evaluated:

1) Are mediums or controls able to detect the cause of death in photographs of deceased individuals at above-chance levels?
2) Are mediums better at this task than controls?
3) Are there any brain activity markers that accurately differentiate between correct and incorrect classification, as well as medium and control participants?

2. Methods

2.1. Participants

Based on previous effects observed in 12 participants (Delorme et al., 2018), twenty-four participants were recruited (12 per group). Twelve professional mediums (who claimed the capacity to connect with deceased individuals based on examination of facial photographs alone) and 12 age-, gender-, and ethnicity-matched controls that claimed to not have that ability (53 ± 9 years; 4 males and 20 females). Mediums were selected from a pool of candidates in the San Francisco Bay Area through word of mouth or from the internet (e.g. https://www.yelp.com/). To be selected as a medium, the individual: (1) had to regularly provide professional "readings" for clients and; (2) be vetted by at least two individuals who claimed that the mediums provided accurate information not readily available to them. For vetting, we often used Yelp reviews and only chose mediums listed there who had obtained dozens of positive reviews. Control participants were recruited through a Craigslist advertisement. Supplemental material 1 and 2 provide more details about the message sent to mediums and control participants.

The inclusion/exclusion criteria for all participants were: Normal or corrected to normal vision; able to sit comfortably in front of a computer screen in a dark room; not currently diagnosed with any psychiatric condition or following any psychoactive drug treatment; able to perform the task; not currently on medications that might have affected EEG and ECG measurements (e.g., beta-blockers or calcium channel blockers; Goodnick, Jerry, & Parra, 2002; Olgin & Zipes, 2007); and able to commute to the IONS laboratory where the experiment was conducted.

Candidates completed an online survey and were then contacted by a Research Assistant by email or phone to confirm inclusion/exclusion criteria requirements. Eligible participants were scheduled for the laboratory session and received session preparation instructions (e.g. instructions to wash their hair prior to the session to ensure EEG signal quality, etc.). The research assistant conducted consenting procedures and all participants signed an informed consent. The participants received a \$100 gift card for their participation. All study activities were reviewed and approved by the Institute of Noetic Sciences Institutional Review Board (approval DELA_2016_01).

2.2. Stimuli

Participants were instructed before the experiment began that the images were balanced for a number of features (detailed later in this report) and were asked to use their intuition to respond. The task involved the presentation of 191 photographs: 11 practice trials with 4 images each of the individuals who died by firearm or car crash, and three images of individuals who died of a heart attack. This practice ensured participants were familiar with the task and response keys before starting the experiment. The experimental task included 60 pictures for each of the three causes of death (i.e. 180 images in total). The images were displayed one at a time on a computer screen for up to 30 s and disappeared after the participants responded by pressing one of three keys on a keypad to indicate the cause of death. The response keys of 1, 2, and 3 corresponding with the cause of death were randomized across participants. After the participants' responses, there was a 3-second delay before the next photo was presented. During the intertrial delay, instructions and key button information were again displayed and participants were instructed to gaze at a centrally located fixation cross. Trial-by-trial accuracy feedback was not provided to the participants. Participants were not told where the images were obtained.

Immediately prior to starting the task, each participant had 11 practice trials with facial images not used in the subsequent task, and practice trial data were not used in subsequent analyses. Image presentation was controlled by the Matlab Psychophysics Toolbox (Brainard, 1997). The size of each image was 320×480 pixels (obtained by doubling the image's original resolution, as explained below) presented at a screen resolution of 800×600 on a CRT (cathode ray tube) monitor to ensure proper timing of the display of each image with millisecond accuracy. The experiment was conducted inside an 8×8 foot double-walled, solid steel, electromagnetically shielded, and electrically-grounded chamber.

All 191 images in the task originated from the "Officer Down Memorial Page" website: https://www.odmp.org/. More than 23,000 police officers who died in the exercise of their duty in the United States from 1791 to 2019 are acknowledged on this website, often with photographs. The study team made efforts to respect the privacy of the individuals depicted and their family members.

2.3. Image selection

The image selection process was designed to minimize the chance of obtaining conventional information that might have been relevant to the classification task. For example, overweight individuals might have been more likely to suffer from cardiac arrest. Thus, to form a balanced image pool – 2,285 available photographs and information associated with them (officer's age when they died, the cause of death, and their photograph) were first downloaded. The four leading causes of death were by firearm (1357), car crash (380), motorcycle crash (166) and heart attack (113). Motorcycle crashes were excluded because they were too similar to car crashes, resulting in three cause-of-death categories. Images where the officer was wearing a hat, or included more than a portrait of the head, or were of poor quality, were excluded by experimenter AD. To balance race across categories, images where the individuals were not obviously Caucasian, African American, or Asian were also removed by experimenter AD. Death by heart attack usually involved older individuals, so to help balance apparent age across the three categories, images depicting younger individuals in the death by firearm category, and some images in death by car crash were randomly removed. This process resulted in 381 photographs: 211 involving death by firearm, 102 by car crash, and 68 by heart attack.

The images were then processed as follows:

- Cropped by manually indicating the position of the left ear, the right ear, the top of the head, and the bottom of the chin (custom software we wrote also used in Delorme et al., 2018);
- Resized to 160 × 240 pixels using linear interpolation while preserving the aspect ratio of the original image (custom software we wrote also used in Delorme et al., 2018).
- Removed background by setting it to transparent using Gimp software (version 2.10; The GIMP Development Team)
- Converted to grayscale by a custom Matlab script using the luminosity method (i.e. using the weighted sum of 0.21*Red + 0.72*Green + 0.07*Blue);
- Normalized by setting the grey level mean for each picture to 122 on a scale of black (or 0) to white (or 255), and setting the standard deviation for the grey level pixels to 55 (on a scale of 0 to 255). Only non-transparent pixels were used to perform this operation, with values below 0 were capped at 0, and values above 255 were capped at 255.

Then each of the 381 processed photographs was independently rated on these eight characteristics by three judges using a program specifically developed for this task: gender (M/F); perceived age (20–30, 30–40, 40–50, 50–60, 60–70); direct gaze (yes/no); glasses (yes/no); head position (facing camera, tilted, profile); smile (yes, no); hair color (light, dark); weight (normal weight, overweight, obese); race (Caucasian, African American, Asian); and picture resolution (acceptable, medium, poor). The first author (AD) explained to each judge prior to scoring the meaning of each category, using examples that were not in the experimental pool. For example, even a slight smile should be considered a smile. Image order was randomized and raters were blind to the cause of death. The median of the three judges' ratings were then used to generate three photo subgroups with 60 images in each cause-of-death category, minimizing the differences between groups on the eight judged characteristics. In addition, the groups were balanced on two low and high spatial-frequency continuous variables which were calculated by average spectral amplitude of the pixel closest to (low) and furthest away from (high) the origin in the 2-D FFT decomposition (the Matlab code used to perform this is available from the first author upon request).

Appropriate statistical tests were then conducted to ensure subgroup similarity (e.g., Fisher test for categorical variables with two values; Chi^2 test for categorical variables with more than 2 values; two-tailed paired *t*-test for spatial-frequency continuous variables). A *p*-value larger than 0.4 (Delorme et al., 2018) was required for all characteristics for all possible pairings of cause-of-death categories (heart attack and firearm; firearm and car crash; car crash and heart attack). This procedure resulted in a total of 180 photographs, with 60 in each cause-of-death category. In addition, 11 images were added for practice trials.

2.4. Machine learning classification

Because the image selection procedure did not remove all possible information related to a combination of features and to further ensure the subgroup balance, a machine learning algorithm was used to see if it could differentiate between the three categories based solely on the 10 characteristics described above.

Three types of classifiers were used: random forest; logistic regression; and support vector machine using functions from the Matlab statistics toolbox. The Matlab *treebagger* function (Matlab Statistics and Machine Learning Toolbox Version 11.5, The Mathworks, Inc.) with 1,000 learners was used for the random forest classifier. The performance was estimated using "out-of-bag classification error," and the out-of-bag matrix was bootstrapped before computing classification performance to form a 95% confidence interval. The Matlab *glmfit* function (with 'logit' link function and binomial distribution) was used for logistic regression, and the Matlab *fitcsvm* function with its default parameters was used for the support vector machine algorithm.

A 10x10 cross-validation procedure was used for both the logistic regression and support vector machine function. This means that the set of 180 images was randomly divided into 10 sets of about 18 images each. Ten iterations were performed in which nine of these sets were used to train the classifier and the remaining set was used to test it. The 95% confidence intervals of the accuracy of classification were calculated for each classifier. None of the classifiers were able to classify images above chance expectations (Table 1). We note that some classifiers resulted in performance significantly below expectations. It is not possible to simply invert the results of those classifiers to obtain significant positive performance because that would entail tailoring the set of classifiers to the specific training/testing dataset, and the results would therefore not generalize. Note that, instead of performing pairwise classifications between the 3 categories of images, we could have used a 3-category Softmax machine learning (ML) solution. However, even when using Softmax 3-class classifiers, an analysis of which category might be biased would require pairwise comparison between classes - which is what we present here.

2.5. EEG, ECG, and behavioral data acquisition

A 64-channel ActiveTwo EEG system (Biosemi, Inc.) with integrated electrocardiography (ECG) measures was used to collect EEG data at a 1,024 Hz sample rate. Electrodes were placed according to the 10–20 nomenclature (standard 64-channel EasyCap). Three sizes of caps were used to accommodate subjects with different head sizes. SignaGel electrode gel was applied to each electrode and active electrode offsets were kept below manufacturer guidelines (Biosemi offset of ± 20). Two auxiliary ECG Biosemi electrodes, positioned under the collar bone (left and right), were used to record participants' heart rate. Signals recorded from the left electrode were subtracted from the right electrode for post-processing.

Behavioral data (the participant's response to each photo) was saved in two ways. First, keypresses were sent to the EEG amplifier digital input channel using the Biosemi USB interface and saved along with the EEG data. One set of markers represented the cause of death category and other markers represented the participant's responses. Second, the latencies of responses were saved in a separate text file on the computer used to control the presentation of the photographs (the presentation computer was different from the computer used to collect the EEG data). We verified that the correspondence of the two data streams on the two computers was within millisecond precision.

2.6. Behavioral data analysis

For each photograph, responses were encoded as being correct or incorrect. Data from images that participants failed to respond to within 30 s were excluded (average misses 2.1 ± 3.4 images; maximum −13 out of 180 images). A general linear model (GLM) with the following variables was used: response (correct vs. incorrect); group (medium vs. control); and type of death. Two regression GLMs were also conducted: 1) percentage response correct on each category above chance expectation as the dependent variable; and 2) reaction time as the dependent variable. The percentage of correct responses was calculated for each classification category for each participant. For example, Participant A

Table 1
Classifier performance for pairs of categories. The percentage range indicates 95% confidence intervals. Bold highlighted cells indicate cases where the classifier was significantly below chance expectation.

	Random Forest	Logistic regression	Support vector machine
Auto vs Heart	38.9%–57.1%	41.5%–50.9%	41.7%–56.7%
Auto vs Gun	40.3%–58.5%	**33.3%–45.6%**	33.3%–50.0%
Heart vs Gun	**26.4%–43.3%**	**37.5%–48.3%**	**34.2%–49.2%**

could have 80% correct on the firearm category, 75% correct for heart attack and 35% correct for vehicle accidents. This allowed for performance comparison across categories.

To evaluate if performance on each category was above chance expectation (33% of correct responses), chance expectation (1/3) was subtracted from each performance percentage (e.g. 33% would be subtracted from 80% correct for firearm in the above example, resulting in 47% above chance expectation for that participant). In this way, the deviation from chance expectation for each category was captured in the intercept of the GLM analysis. Percentage data passed the Kolmogorov Smirnov (KS) tested for normality (p >0.2). Reaction time was not normally distributed so a log transform was used. The GLM residuals using this transformation were acceptably close to being normally distributed (KS normality test, $p = 0.1$).

2.7. ECG data processing

ECG data was extracted from the Biosemi BDF files and filtered (high pass = 0.05 Hz; low pass = 100 Hz; Bailey et al., 1990) using EEGLAB software 2019.0 (Delorme & Makeig, 2004). Data was then divided by 32 in Matlab to convert the 24-bit signal to microVolts (ref: https://www.biosemi.com/activetwo_full_specs.htm). ECG time series were saved as a csv file, which was then imported into Kubios HRV Premium v 3.1.0 (University of Kuopio, Kuopio, Finland) to generate R-R intervals, heart rate, standard deviation of normal RR intervals (SDNN), low (LF), and high (HF) frequency domain measures (LF: 0.04–0.15 Hz; HF: 0.15–0.4 Hz). Heart rate variability (HRV) analysis parameters included a 100-second window width, 50% window overlap; autoregressive spectrum model order = 16 with no factorization, and interpolation rate of 4 Hz.

2.8. EEG data analysis

EEG data were imported into EEGLAB software 2019.0 (Delorme & Makeig, 2004) running in Matlab R2018b. Raw data were downsampled to 512 Hz, detrended, and filtered using an FIR filter at 1 Hz (non-causal zero-phase distortion highpass filter of length 1691 samples with transition bandwidth of 1 Hz, passband edge of 1 Hz and cutoff frequency (−6 dB) of 0.5 Hz) and lowpass filtered at 55 Hz (noncausal zero-phase distortion 125 points lowpass filter with transition bandwidth 13.75 Hz), passband edge of 55 Hz and cutoff frequency (−6 dB) of 61.9 Hz). Defective EEG channels were identified manually and removed in each participant (2.5 ± 2.0 channels removed on average) and interpolated using spherical splines for group analysis (Perrin et al., 1989). Data sections containing obvious artifacts (e.g. body movements, jaw clenching, etc.), were removed by visual inspection of the filtered data (2 ± 2 min of data out of on average 20-minute recordings). All artifact removal was performed blind with respect to the behavioral responses.

Data were then average-referenced using the EEGLAB function *pop_reref*, temporarily interpolating removed channels for performing average reference. Infomax Independent Component Analysis (ICA) was performed to separate ocular and muscular artifacts (i.e. eye blinks and lateral eye movements) using the ICLabel v.1.1 EEGLAB plugin (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) to classify components and to automatically select those components which were likely to be eye or muscle artifacts (likelihood superior to 90%). Finally, single-trial event-related potentials (ERP) from −0.3 s to +1 s post-image presentation were extracted for each type of face photography (no baseline was removed).

Plotting the RMS (Root Mean Square) of all of the electrodes across time (Fig. 1) led to manual selection of two peaks of interest: *P*80-110 (i.e. from 80 to 110 ms) and *P*200-350. Note that this procedure was performed blindly with respect to experimental conditions because all conditions and subject groups were pooled.

For each subject and each peak of interest, a first level General Linear Model (GLM) with weighted least square optimization was used – linear modeling plugin (LIMO) version 2.0 extension of the EEGLAB software (Pernet, Chauveau, Gaspar, & Rousselet, 2011). The type of response (correct and incorrect) and the cause of death were used as factors in an interaction design. Six beta parameters (correct-heart, correct-auto, correct-firearm, incorrect-heart, incorrect-auto, incorrect-firearm) were obtained. These beta parameters were then averaged over the time range of interest (80–110 ms and 200–350 ms as defined above), exported as a text file, and fed into a second-level GLM to assess random effects across subjects using Statistica 13.0 (TIBCO Software Inc.). This type of hierarchical analysis is standard in brain imaging analysis (Stephan, Mattout, David, & Friston, 2006). Uncorrected parametric statistics were adjusted for multiple analyses using the False Discovery Rate algorithm (FDR; Benjamini & Yekutieli, 2001). Electrodes significant at $p < 0.01$ after FDR are shown as the larger black disks in Fig. 2.

3. Results

3.1. Behavioral results

Of the three research questions investigated, the first two were tested using behavioral responses. The first question was whether mediums or controls would be able to detect the cause of death in facial images of deceased individuals. When considering all participants, there was a significant main effect of response accuracy ($F(1,66) = 8.77$; partial $\eta^2 = 0.12$; $p = 0.004$).

The second question was whether mediums were better at this task than controls. Here we observed a main effect of group ($F(1,66) = 8.5$; partial $\eta^2 = 0.11$; $p = 0.005$), with controls more accurate than mediums (accuracy of 4.0% above chance expectation for controls, as compared to −0.2% for mediums). Accuracy levels were not different across the three

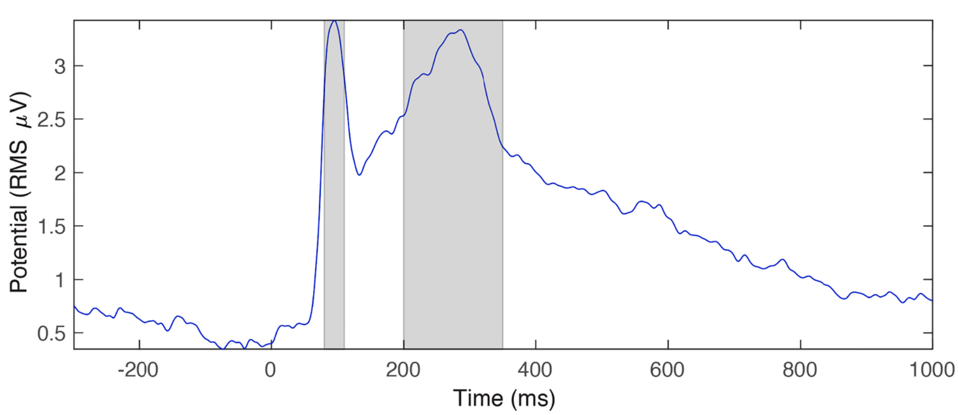

Fig. 1. Regions of interest selected in the root mean Square (RMS) event-related potential (ERP). We manually selected two regions, between 80 and 110 ms and between 200 and 350 ms.

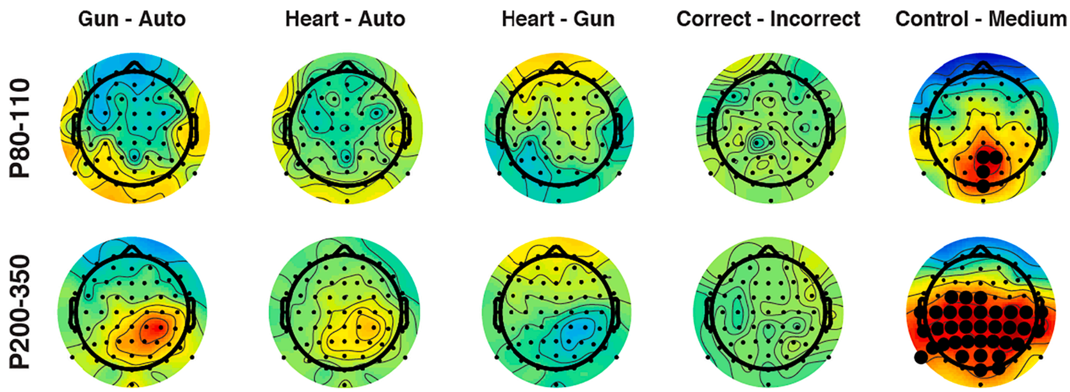

Fig. 2. Partial correlation coefficients and significance for each categorical variable in the GLM model. Interactions between variables are not shown but none indicated significant effects. The top row represents the average activity between 80 and 110 ms and the bottom row represents the average activity between 200 and 350 ms. The black dots show significant electrodes at 0.01 threshold after correction for multiple comparisons using FDR.

causes of death ($F(2,66) = 0.52$; $p = 0.59$).

For the reaction time measure, we also observed a clear difference between mediums and controls. A group main effect ($F(1,66) = 221$; $\eta^2 = 0.05$; $p <$ 1e-10) showed that mediums took on average 4911 ms (± 87 ms) to answer, compared to 3177 ms (± 61 ms) for controls. An interaction was also observed between response type (correct vs. incorrect) × group ($F(1,66) = 4.31$; $\eta^2 = 0.001$; $p = 0.04$), with mediums being slower to respond to images when their responses were incorrect vs. correct (5023 ms vs. 4686 ms), while controls showed the opposite trend (3132 ms vs 3251 ms). There was also an interaction between group, response type, and image type ($F(1,66) = 3.64$; $\eta^2 = 0.002$; $p = 0.03$). No other effects or interactions were observed.

3.2. EEG results

The third research question was whether physiological measures would discriminate between correct vs. incorrect assignments of the causes of death, as well as medium vs. control participants. To examine this question, first, the time periods of interest in the EEG analysis were reduced to 2 peaks of interest (Fig. 1) to decrease the number of statistical tests performed (see Methods).

Robust differences between controls and mediums were observed mostly in the occipital regions for both *P*80-110 and *P*200-350. No other differences or interactions were obtained. Although we observed differences between groups, we did not observe any brain activity markers that reliably differentiated between correct and incorrect classification.

3.3. Heart rate results

Mean heart rate was significantly higher ($t = 2.39$; $df = 22$; $p = 0.03$) for mediums (76 bpm) than controls (67 bpm). Differences in heart rate variability measures failed to reach significance: SDNN ($t = 0.92$; $df = 22$; $p = 0.37$), LF HRV ($t = 0.89$; $df = 22$; $p = 0.38$) and HF HRV ($t = 0.75$; $df = 22$; $p = 0.45$).

4. Discussion

Three hypotheses were evaluated through a classification task where mediums and controls selected the cause of death of deceased persons by viewing facial images of those individuals. The EEG and ECG of the experimental participants were recorded while they performed this task. Hypothesis number one was demonstrated as, overall, participants were able to detect the cause of death of deceased individuals at statistically robust above-chance levels. Contrary to our expectations that mediums would perform better than controls, the controls performed significantly better than the mediums. Third, there was no overall evidence for the presence of EEG markers corresponding to accurate classification of the images, although different EEG and ECG markers between the mediums and controls were observed.

4.1. Limitations

It is possible that participants were able to differentiate individuals who died of disease from those who died from accidents by taking advantage of subtle facial cues in the images. For example, facial features have been shown to indicate signs of cardiovascular disease (Mette et al., 2014) and cigarette smoking (Okada, Alleyne, Varghai, Kinder, & Guyuron, 2013). Features of adolescents' faces may also predict adult health and mortality (Reither, Hauser, & Swallen, 2009). The use of subtle cues cannot be ruled out in the current experiment. However, we think it is unlikely because the machine learning classifier failed to classify images above chance expectations. It is also unlikely because there were no observed differences in accuracy of mediums and controls among the three categories. Had subtle health cues biased the results, then the heart attack category should have had increased accuracy as compared to death by firearm or accident, and this was not observed. In a future version of this experiment, a separate group of participants could be probed about their subjective assessment of the images (e.g., on factors like impulsiveness or health). With those assessments in hand, it might be possible to help reduce or partial out such biases from the image categorization task.

Contrary to our second hypothesis, controls were found to more accurately discern the cause of death than the mediums. There are limitations to recruiting mediums through Yelp as some mediums might have obtained fake reviews, but Yelp's methods for blocking fake reviews help to counteract that possibility. Differences in motivation probably cannot explain such results, as one would expect mediums to be more motivated than controls. On the other hand, greater motivation might have produced performance anxiety, which could be a potential factor. The Research Assistant noted that some mediums expressed concern about their performance and were very interested in learning how well they did. Mediums had significantly higher heart rate during the task itself, which supports the conjecture that they felt more pressure to succeed which, in turn, may have interfered with their performance. However, on average mediums may simply have higher baseline heart rate levels than controls (Ulrich-Lai & Herman, 2009), which could be tested in a follow-up study.

Controls had faster response speeds (by approximately 2 s on average) and better performance, which may indicate that they did not reflect on their answer as long as mediums did. Improved performance for similar intuitive tasks has been observed for faster vs. slower responses (Cardeña, 2018; Kahneman, 2013). If intuition relies on emotions and perceptions arising before thoughts are formulated, then the mediums' slower responses may have impaired their accuracy (Volz &

Yves von Cramon, 2006). For example, high scores on the Cognitive Reflection Test (which underlies analytical thinking) is correlated with lower scores on divergent thinking (Corgnet, Espín, & Hernán-González, 2016). Regardless of whether participants were mediums or control, this raises the question of whether faster responses showed improved performance. To explore this, a regression analysis was conducted between performance and median reaction time across all subjects. Although the slope was negative, the result was not statistically significant ($r^2 = 0.015$; $p = 0.56$).

If not due to chance fluctuations, then how might we explain the difference in performance between mediums and controls? Some of the mediums commented that they found it difficult to differentiate between the type of death, as they reported feeling the pain of the deceased individual, but not the cause of that pain. They might have interpreted a given type of pain as a heart attack, but a similar pain could have occurred by being shot in the chest, or by chest trauma associated with a car accident. Given this feedback from the mediums, future experiments involving cause-of-death categories should be designed to avoid similar confusions. Additionally, mediums reported that the time pressure of the experiment did not allow them to really connect with the deceased individuals like they normally would, and so they felt forced to use different strategies to try to respond as fast as possible.

4.2. EEG analysis

Our approach to ERP analysis consisted first in identifying the time period of the largest deflection compared to baseline, then analyzing those regions. This approach might ignore changes in activity in a reduced subset of channels. However, due to volume conduction, any change in brain activity tends to recruit a large set of scalp channels, so the most important portions of the grand average ERP were likely captured. The portion between 80 and 110 ms probably corresponds to the visual P100. This early ERP activity, at about 100 ms, is influenced by both visual detail information (Hopf & Mangun, 2000; Taylor, McCarthy, Saliba, & Degiovanni, 1999) and facial configuration of visual stimuli (Halit, de Haan, & Johnson, 2000). Differences in this early visual peak associated with accurate mediumship performance were previously noted (Delorme et al., 2018) with higher potential amplitude associated with decreased performance in right parietal areas, a finding we did not observe in this report. We did, however, note that controls had higher accuracy than mediums, and they also had increased activity in occipital regions. Interestingly, this potential is also influenced by attention (Mangun & Hillyard, 1991). Because it is unlikely that differences between mediums and controls would be due to low-level facial information, we may speculate that this effect reflected different types of allocation of attention.

The later peak at 200 to 350 ms most likely reflects higher sensory processing and corresponds to the N200, which is a negative-going wave that peaks 200–350 ms post-stimulus and is found primarily over anterior scalp sites (Folstein & Van Petten, 2008). The N200 ERP is known to be modulated by conflicting situations like the Eriksen Flanker Task (Heil, Osman, Wiegelmann, Rolke, & Hennighausen, 2000) and go/no-go paradigm (Pfefferbaum, Ford, Weller, & Kopell, 1985) and is thought to reflect response inhibition. Given the ERP amplitude we observed in this latency range was larger for controls as compared to mediums, faster response time and accuracy for controls might be linked to better N200-amplitude mediated stimulus identification (Patel & Azzam, 2005), overcoming stereotypical responses or conflict monitoring (Azizian, Freitas, Parvaz, & Squires, 2006), and detection of novelty or mismatch (Folstein & Van Petten, 2008). This is consistent with the performance anxiety some mediums were reporting, and possibly with their increased heart rate. Assuming the overall effect we observed on image type detection was robust, this could potentially explain the mediums' poor behavioral performance.

5. Conclusion

In summary, participants (as a whole) were able to successfully categorize the type of death above chance expectation, and there were differences observed in EEG activity in how mediums and controls processed the facial photographs. These results warrant further investigation in a larger and more diverse pool of participants.

The anonymized behavioral and EEG data used in this study are available to the scientific community (DOI: https://doi.org//10.5281/zenodo.3600490). The images, presentation and data analysis scripts used in our study are available from the first author upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank the Bial Foundation for generously supporting this study (Grant number 188/16) and Nina Fry for proofreading the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2020.105638.

References

Azizian, A., Freitas, A. L., Parvaz, M. A., & Squires, N. K. (2006). Beware misleading cues: Perceptual similarity modulates the N2/P3 complex. *Psychophysiology, 43*(3), 253–260.

Bailey, J. J., Berson, A. S., Garson, A., Horan, L. G., Macfarlane, P. W., Mortara, D. W., & Zywietz, C. (1990). *"Recommendations for Standardization and Specifications in Automated Electrocardiography: Bandwidth and Digital Signal Processing. A Report for Health Professionals by an Ad Hoc Writing Group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology*. American Heart Association".

Beischel, J., Mosher, C.a., & Boccuzzi, M. (2015). The possible effects on bereavement of assisted after-death communication during readings with psychic mediums: A continuing bonds perspective. *Omega: Journal of Death and Dying, 70*(2), 169–194.

Beischel, J., & Schwartz, G. E. (2007). Anomalous information reception by research mediums demonstrated using a novel triple-blind protocol. *EXPLORE, 3*(1), 23–27.

Beischel, J., & Zingrone, N. L. (2015). Mental Mediumship. In *Parapsychology: A handbook for the 21st century* (pp. 301–313). Jefferson, NC, US: McFarland & Co.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist., 29*(4), 1165–1188.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663–677. https://doi.org/10.1037/amp0000236.

Mette, C., Ruth, F.-S., Peter, S., Jensen, G. B., Nordestgaard, B. G., & Anne, T.-H. (2014). Visible age-related signs and risk of ischemic heart disease in the general population. *Circulation, 129*(9), 990–998.

Corgnet, B., Espín, A. M., & Hernán-González, R. (2016). Creativity and cognitive skills among millennials: Thinking too much and creating too little. *Frontiers in Psychology, 7*.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*, 9–21.

Delorme, A., Beischel, J., Michel, L., Boccuzzi, M., Radin, D., & Mills, P. J. (2013). Electrocortical activity associated with subjective communication with the deceased. *Frontiers in Psychology, 4*.

Delorme, A., Pierce, A., Michel, L., & Radin, D. (2018). Intuitive assessment of mortality based on facial characteristics: Behavioral, electrocortical, and machine learning analyses. *EXPLORE, 14*(4), 262–267.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170.

Goodnick, P. J., Jerry, J., & Parra, F. (2002). Psychotropic drugs and the ECG: Focus on the QTc interval. *Expert Opinion on Pharmacotherapy 11996627, 3*(5), 479–498. https://doi.org/10.1517/14656566.3.5.479.

Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *NeuroReport, 11*(9), 1871–1875.

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-Task: Inhibitory Executive Process? *Journal of Psychophysiology, 14*(4), 218–225.

Hopf, J. M., & Mangun, G. R. (2000). Shifting Visual attention in space: An electrophysiological analysis using high spatial resolution mapping. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 111*(7), 1241–1257.

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st edition). New York: Farrar, Straus and Giroux.

Knapp, K. D. (2017). *William James: Psychical research and the challenge of modernity*. Chapel Hill: The University of North Carolina Press.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology. Human Perception and Performance, 17*(4), 1057–1074.

Okada, H. C., Alleyne, B., Varghai, K., Kinder, K., & Guyuron, B. (2013). Facial changes caused by smoking: A Comparison between smoking and nonsmoking identical twins. *Plastic and Reconstructive Surgery, 132*(5), 1085–1092.

Olgin, J., & Zipes, D. (2007). Specific arrhythmias: Diagnosis and treatment. In P. Libby, R. Bonow, D. Mann, & D. Zipes (Eds.), *Braunwald's heart disease: A textbook of cardiovascular medicine*. WB Saunders.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences*, 147–154.

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: A toolbox for hierarchical linear modeling of ElectroEncephaloGraphic Data. *Computational Intelligence and Neuroscience, 2011*.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol, 72*(2), 184–187.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to Response Production and Inhibition. *Electroencephalography & Clinical Neurophysiology, 60*(5), 423–434.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). The ICLabel Dataset of Electroencephalographic (EEG) Independent Component (IC) Features. *Data in Brief, 25*.

Reither, E. N., Hauser, R. M., & Swallen, K. C. (2009). Predicting Adult health and mortality from adolescent facial characteristics in yearbook photographs. *Demography, 46*(1), 27–41.

Rock, A. J., Beischel, J., & Cott, C. C. (2009). Psi vs. Survival: A qualitative investigation of mediums' phenomenology comparing psychic readings and ostensible communication with the deceased. *Transpersonal Psychology Review, 13*(2), 76–89.

Stephan, K. E., Mattout, J., David, O., & Friston, K. J. (2006). Models of functional neuroimaging data. *Current Medical Imaging Reviews, 2*(1), 15–34.

Taylor, M. J., McCarthy, G., Saliba, E., & Degiovanni, E. (1999). ERP Evidence of developmental changes in processing of faces. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 110*(5), 910–915.

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews. Neuroscience, 10*(6), 397–409.

Volz, K. G., & Yves von Cramon, D. (2006). What neuroscience can tell about intuitive processes in the context of perceptual discovery. *Journal of Cognitive Neuroscience, 18*(12), 2077–2087.