RESEARCH ARTICLE

REVISED **Brain-to-Brain (mind-to-mind) interaction at distance: a confirmatory study [version 3; referees: 1 approved, 1 not approved]**

Patrizio Tressoldi[1], Luciano Pederzoli[2], Marco Bilucaglia[2], Patrizio Caini[2], Pasquale Fedele[3], Alessandro Ferrini[2], Simone Melloni[2], Diana Richeldi[2], Florentina Richeldi[2], Agostino Accardo[4]

[1]Dipartimento di Psicologia Generale, Università di Padova, Padova, 35131, Italy
[2]EVANLAB, Firenze, 50134, Italy
[3]LiquidWeb, Siena, 53100, Italy
[4]Dipartimento di Ingegneria e Architettura, Università di Trieste, Trieste, 34127, Italy

Abstract

This study reports the results of a confirmatory experiment testing the hypothesis that it is possible to detect coincidences of a sequence of events (silence-signal) of different length, by analyzing the EEG activity of two human partners spatially separated when one member of the pair receives the stimulation and the second one is connected only mentally with the first. Seven selected participants with a long friendship and a capacity to maintain focused mental concentration, were divided into two groups located in two different laboratories approximately 190 km apart. Each participant acted both as a "stimulated" and as a "mentally connected" member of the pair for a total of twenty sessions overall.

The offline analysis of EEG activity using a special classification algorithm based on a support vector machine, detected the coincidences in the sequence of events of the stimulation protocol between the EEG activity of the "stimulated" and the "mentally connected" pairs.

Furthermore the correlation of the power spectra of the five EEG frequency bands between each of the twenty pairs of data was analyzed using a bootstrap procedure.

The overall percentage of coincidences out of 88 events was 78.4% and the statistically significant average correlations between the EEG alpha and gamma bands among the pairs of participants, confirmed the results observed in a pilot study. The examination of potential internal, external and statistical artifacts which might have caused these results, ruled out external and internal artifacts. However, the examination of potential statistical artifacts revealed a good level of coincidences in only four pairs using a new procedure to detect the sequences of silence and signal between the EEG activity of the pairs of participants, giving a mild support to the hypothesis that two brains and hence two minds can be connected at distance.

Open Peer Review

Referee Status: ☑ ☒

	Invited Referees	
	1	**2**
REVISED **version 3** published 23 Oct 2014		☒ report
		↑
REVISED **version 2** published 29 Sep 2014		☒ report
		↑
version 1 published 05 Aug 2014	☑ report	☒ report

1 **James Lake**, University of Arizona USA

2 **Sam Schwarzkopf**, University College London UK

Discuss this article

Comments (3)

Corresponding author: Patrizio Tressoldi (patrizio.tressoldi@unipd.it)

Competing interests: No competing interests were disclosed.

REVISED **Amendments from Version 2**

The main revisions to Version 2 are:

- Added control of potential statistical and methodological artifacts to the reported results. In particular, a new method to calculate the coincidences of the sequence of silence and signal events between the EEG activity of pairs of participants has been applied and a test of specificity of the results has been applied both to this new and the correlation results;

- A new Figure 2 presenting the results of the new method to calculate the coincidences;

- The current Figure 3 and Table 2 have been modified;

- An update of the raw data and results available at http:// figshare.com/articles/BBI_Confirmatory/1030617;

- A reply to both reviewers' comments, posted in the comments section of the online version.

See referee reports

Introduction

Brain-to-brain interaction (BBI) at distance, that is, outside the range of the five senses, has been demonstrated by Pais-Vieira *et al.*, (2013), by connecting the brains of rats via an Internet connection.

A similar effect has been demonstrated with humans in a pilot study by Rao & Stocco, (2013) by sending the EEG activity generated by a subject imagining to move his right hand via the Internet to the brain of a distant partner which triggered his motor cortex causing the right hand to press a key. Similarly, Grau *et al.*, (2014) were able to induce the conscious perception of light flashes to a participant, triggering a robotized transcranial magnetic stimulation by a signal generated from the EEG correlates of the voluntary motor imagery from a partner located 5,000 miles apart and transmitted via the Internet.

Even though there is cultural resistance in accepting the possibility of observing similar effects in humans without an internet connection, some evidence of these effects nevertheless exists. A comprehensive search of all studies related to this line of research has revealed at least eighteen studies from 1974 until the present time (see Supplementary Material).

In all these studies the principal aim was to observe whether the brain activity evoked by a stimulus (e.g. by presenting light flashes or images) in one member of a couple, could also be observed in the brain of the partner. Even if some of these studies, those using functional neuroimaging, can be criticized for potential methodological weaknesses that could account for the reported effects (Acunzo *et al.*, 2013), the questions is still open as to whether or not it is possible to connect two human brains at distance.

The possibility of connecting the brains of two humans at distance without using any classical means of transmission is theoretically expected if it is assumed that two brains, and consequently two minds, can be entangled in a quantum-like manner. In quantum physics, entanglement is a physical phenomenon that occurs when pairs (or groups) of particles interact in ways such that the measurement (observation) of the quantum state (e.g. spin state) of each member is correlated with the others, irrespective of their distance without apparent classical communication.

At present, generalizability from physics variables to biological and mental variables can be done only by analogy given the differences in their properties, but some theoretical models are already available. For example in the Generalized Quantum Theory (Filk & Römer, 2011; Von Lucadou & Romer, 2007; Walach & von Stillfried, 2011), "entanglement can be expected to occur if descriptions of the system that pertain to the whole system are complementary to descriptions of parts of the system. In this case the individual elements within the system, that are described by variables complementary to the variable describing the whole system, are non-locally correlated".

Reasoning by analogy, we hypothesized the possibility of entangling two minds, and consequently two brains as complementary parts of a single system and studying their interactions at distance without any classical connections.

In a pilot study, Tressoldi *et al.*, (2014) tested five couples of participants with a long friendship and a capacity to maintain a focused mental concentration, who were separated by a distance of approximately five meters without any sensorial contact. Three sequences of silence-signal events lasting two and half minutes and one minute, respectively, were delivered to the first member of the pair. The second member of the pair was simply requested to connect mentally with his/her partner. A total of fifteen pairs of data were analyzed. By using a special classification algorithm, these authors observed an overall percentage of correct coincidences of 78%, ranging from 100% for the first two segments silence-signal, to approximately 43% in the last two. The percentages of coincidences in the first five segments of the protocol were above 80%. Furthermore a robust statistically significant correlation was observed in all but beta EEG frequency bands, but was much stronger in the alpha band.

These preliminary results of the pilot study prompted us to devise this pre-registered replication study.

Methods

Study pre-registration

In line with the recommendations to distinguish exploratory versus confirmatory experiments (Nosek, 2012; Wagenmakers *et al.*, 2012), we pre-registered this study in the Open Science Framework site (https://osf.io/u3yce).

Participants

Seven healthy adults, five males and two females, were selected for this experiment and included as co-author. Their mean age was 41.7, SD = 16.6. Inclusion criteria were their friendship lasting more than five years and their experience in maintaining a focused mental concentration resulting from their experience (ranging from four to fifteen years) in meditation and other practices to control mental activity, e.g. martial arts practices, yoga, etc.

Ethics statement

Participation inclusion followed the ethical guidelines in accordance with the Helsinki Declaration and the study was approved by the Ethics Committee of Dipartimento di Psicologia Generale, prot.n.63, 2012, the institution of the main author. Before taking part in the experiment, each participant provided written consent after reading a brief description of the experiment.

Apparatus

Ad-hoc software written in C++ for Windows 7, designed by one of the co-authors, SM, controlled the delivering of the choice of the protocols of stimulation and the timing of the EEG activity recordings of the two partners. EEG activity was recorded by using two Emotiv® EEG Neuroheadsets connected wirelessly to two personal computers running Windows 7 OS and synchronized with the atomic clock. The Emotiv® EEG Neuroheadset technical characteristics are 14 EEG channels based on the International 10–20 locations (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, plus 2 references), one mastoid (M1) sensor acts as a ground reference point to which the voltage of all other sensors is compared. The other mastoid (M2) is a feed-forward reference that reduces external electrical interference. Sampling rate is 128 Hz, bandwidth 0.2–45 Hz with digital notch filters at 50 and 60 Hz. Filtering is made by a build in digital 5th order sinc filter and connectivity is obtained by a proprietary wireless connection at the 2.4 GHz band.

Stimuli

One auditory clip was delivered binaurally at a high volume (80 dBs) to one of the partners through Parrot ZIK® headphones connected with the PC controlling the stimulus delivery and EEG recordings. This clip, reproducing a baby crying, was selected among the list of the worst sounds (Cox, 2008) in order to enhance the EEG activity of the stimulated person.

Stimulation protocol

In contrast to the pilot study, the stimulation protocol consisted of three different sequences of 30 seconds of listening to the auditory clip interspersed by silent periods lasting one minute (in the pilot study the durations were twice this length). The three sequences comprised 3, 5 and 7 segments (i.e. silence-signal-silence-signal-silence-signal-silence) and were selected by a random algorithm using the rand function of C++ (in the pilot study only a sequence of 7 segments was used). To prevent any possible prediction of the start of the sequence of events, its presentation was randomly delayed by 1, 2 or 3 minutes.

Procedure

We devised a procedure aimed at recreating a real situation when there is an important event to share, in this case a communication relating to a baby crying. In order to isolate the two partners, four of them were located in a laboratory of the Department of General Psychology of Padova University and the remaining three were placed in the EvanLab a private laboratory located in Florence, approximately 190 km away. A research assistant was present at each location.

The partner designated as "sender" received the following instructions: "*when ready, you must concentrate in silence for one to three minutes to relax and prepare to receive the stimulation to send to your partner. To facilitate your mental connection with him/her, you will see a photo of his/her face via the special glasses* (virtual glasses model Kingshop OV2, see Figure S1 in the Supplementary Material). *Your only task is to endeavor to send him/her mentally what you will hear, reducing your body and head movements in order to reduce artifacts. You will hear a sequence of a baby crying lasting 30 seconds, separated by one minute intervals. The experiment will last approximately 10 minutes*".

The instructions to the second partner designated as "receiver" were: "*when ready, you must concentrate in silence for one to three minutes to relax and prepare to receive the stimulation sent by your partner. To facilitate your mental connection with him/her, you will see a photo of his/her face via the special glasses. Your task is to connect with him/her mentally attempting to receive the stimulation he/she is hearing, reducing your body and head movements in order to reduce artifacts. The experiment will last approximately 10 minutes*".

When all devices were set up, the "sender" was continuously presented with the image of the "receiver" except when the signal was delivered. In this case, the image of a baby crying associated with the auditory clip, replaced the previous one. On the contrary, the "receiver" was continuously presented with only the image of the "sender" up to the end of the session without any further auditory and visual cues that could inform him/her about what the "sender" perceived and listened.

After both partners gave their approval to begin the experiment, the main research assistant located in the EvanLab, started the experiment by informing the second research assistant connected via the Internet to trigger the software controlling the experiment. At the end of the experiment, both partners were informed that it was over. After a break, the partners reversed their roles if available.

Pairing each participant located in one laboratory with each participant located in the second laboratory, a total of 22 pairs of data were collected, because two participants contributed to only three sessions. Two pairs of data were eliminated due to a faulty recording of the EEG activity.

Data analysis
Classification algorithm

The BrainScanner™ classification software was originally developed and is available from one of the co-authors P.F. (Pasquale Fedele p.fedele@liquidweb.it). The analysis was carried out offline separately for each pair taking as input the raw data recorded by the two Emotiv® EEG Neuroheadsets using the procedure and parameters which yielded the best classification accuracy in the pilot study. The first analysis was a classical principal component analysis (PCA) to reduce the data obtained by the fourteen channels to their latent variables. Fifty percent of these data, randomly sampled related to all signal and silence segments were used for the training of the C-supported vector classification (C-SVC) machine (Chang & Lin, 2011; Steinwart & Christmann, 2008).

Regarding the kernel choice, the one that gave the best performance during the pilot tests was the RBF (radial basis function). A general

description of the Supported vector machines (SVMs) is reported in the Supplementary Material.
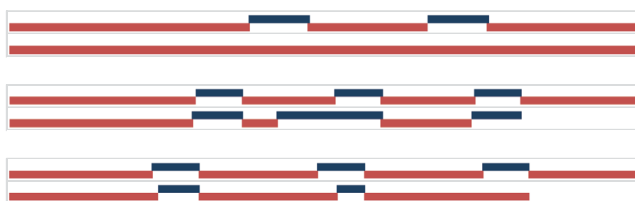
After the training phase, the algorithm was ready to generalize the obtained classification model to all the data, matching the sequence of events of the stimulation protocol with the whole EEG activity. The result was a contingency table (see examples in Figure 1). To control the reliability of these results, the whole procedure was repeated five times and the results were identical.

From the contingency table of each participant with the role of "receiver" we counted the number of coincidences and the number of errors and missing. Given our interest in detecting the sequence of binary events (silence-signal) and not their absolute overlap, a signal detected in the EEG activity of the receiver was considered as a coincidence if at least one of its boundaries (initial or final) overlapped with that of the sender (see examples in Figure 1).

To check the reliability of the scoring system, the data were analyzed independently by two co-authors, PT and SM. Their overall agreement was 89.3%; discrepancies were solved re-checking the original data. All the individual raw data and results are available for independent analyses at http://figshare.com/articles/BBI_Confirmatory/1030617.

Correlational analyses

To have convergent evidence of the relationship between the EEG activity of the two partners, we correlated their EEG activity related to the signal and silence periods recorded in the fourteen channels, with respect to the five frequency bands, delta, theta, alpha, beta and gamma normalized with respect to the total power. Each period of silence and stimulation was divided into tracts of 4 seconds and the Power Spectral Density (PSD) was computed by the periodogram method. The five spectral bands were distinguished as follows: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta

(15–30 Hz) and gamma (30–45 Hz). The PSD of the different bands collapsing all silence and signal periods, was then expressed in normalized units by dividing the power in each band by the sum of the powers in all the bands.

To test the significance of the correlation coefficient we adopted a distribution-free approach, the bivariate non-parametric bootstrap (Bishara & Hittner, 2012) with 5000 iterations. From the sampling distribution, we computed the 95% confidence interval following the percentile method. The bivariate test rejects the null hypothesis if $r = 0$ is not included within the confidence interval. The overall results are reported in Table 2 whereas the results of each of the 20 pairs are reported in Supplementary Table S1. The raw data and the software source code in MatLab "Accardo_Confirmatory_rev.m" are available at http://figshare.com/articles/BBI_Confirmatory/1030617.

Results
Statistical approach

Instead of a traditional Null Hypothesis Significant Testing, we adopted a frequentist parameter estimation approach in line with the APA (2008) and the statistical reform recommendation (Cumming, 2014) and a Bayesian models comparison approach as suggested by Wagenmakers *et al.*, (2011).

Coincidences

The numbers of coincidences in the EEG activity of the participants with the role of "receiver" (the data of the participants with the role of "sender" are irrelevant in this case) detected by the BrainScanner™ classifier, related to the three different stimulation protocols in the twenty sessions are reported in Table 1a, Table 1b and Table 1c. The expected number of coincidences is zero. A percentage of coincidences of the silence and signal events well above the number of missing values and errors, can be a demonstration of a brain (mind) connections between the pairs of participants unless statistical or procedural artifacts can explain them.

Figure 1. Three examples of the matrices of coincidence between the protocol of stimulation and the EEG activity recorded in the "receiver" brain. The first row of each example shows the timing and the sequence of periods of silence and stimulation as delivered to the "sender" brain. The first row of each example shows the timing and the sequence of periods of silence and stimulation as delivered to the "sender" brain. The second row of each example shows the timing and the sequence of the periods of silence and stimulation identified by the BrainScanner™ classifier in the "receiver" brain. Red color = silence; Black color = signal. The first example represents what it is expected if there was no mental connection, the second and the third one represents what we would have observed with mental connection. Using the criteria to consider a coincidence a segment of the protocol with at least one timing boundary (initial or final) overlapped between the two rows, in the second example we count 6 coincidences and 1 omission and in the third example 5 coincidences and 2 omissions.

Table 1a. Coincidences in the EEG activity of the "receivers" detected by the BrainScanner™ classifier, related to the first protocol.

No. 9	Silence	Signal	Silence	% Detection Accuracy
Silence	9			100
Signal		9		100
Silence			9	100

Table 1b. Coincidences in the EEG activity of the "receivers" detected by the BrainScanner™ classifier, related to the second protocol.

No. 8	Silence	Signal	Silence	Signal	Silence	% Detection Accuracy
Silence	8					100
Signal		8				100
Silence			7			87.5
Signal				3		37.5
Silence					2	25

Table 1c. Coincidences in the EEG activity of the "receivers" detected by the BrainScanner™ classifier, related to the third protocol.

No. 3	Silence	Signal	Silence	Signal	Silence	Signal	Silence	% Detection Accuracy
Silence	3							100
Signal		3						100
Silence			3					100
Signal				2				66.7
Silence					2			66.7
Signal						1		33.3
Silence							0	0

The overall percentages of coincidences and their precision were estimated with the corresponding confidence intervals. The classification algorithm correctly detected 69/88 = 78.4%; 95% CI: 68.7–85.7 events, 26/34 = 76.4%; 95% CI: 58.4–87.5 related to the signals and 43/54 = 79.6%; 95% CI: 67.1–88.2 related to the silence events.

Furthermore the Bayes Factor comparing the hypothesis that the percentage of coincidences will outperform the percentage of errors and missing data with the hypothesis of null difference between these two percentages, was calculated with the online applet available at http://pcl.missouri.edu/bf-binomial, using a uniform prior probability distribution based on a beta distribution.

The corresponding Bayes Factors comparing the H1 (above chance detection) vs H0 (chance detection) hypothesis, for the overall and the signal coincidences are 390,625 and 27.1 respectively.

It is interesting to observe that for all three stimulation protocols, the percentages of coincidences of the first three events (silence-signal-silence) was 98.3%, dropping to 40.9% for the next two events (signal-silence) and to 16.6% for the last two events (signal-silence). This drop was also observed in the pilot study, even if it was less dramatic: 83.3% and 43.3%, respectively. However it is important to recall that in the pilot study, the duration of the signals and the silence periods were 60 seconds and 180 seconds, respectively. A plausible explanation of this difference can be the limitation of the present version of our classifier to extract sufficient information to differentiate the two classes of events from the

EEG activity, postulating that the signal/noise ratio of EEG activity reduced after a sequence of three events.

Control of potential statistical artifacts

To control if the above results might represent true coincidences between the EEG activity of the pairs of participants or simply the results of the BrainScanner™ efficiency in detecting sequences of differential EEG activity in the participants with the role of "receivers", we redid all comparisons by training the BrainScanner™ with the EEG activity of the participants with the role of "senders" generalizing the obtained classification model to the EEG activity of the participants with the role of "receivers". The outputs of this new comparison are available in the file NewCoincidences at http://figshare.com/articles/BBI_Confirmatory/1030617.

Only five pairs, 8,11,13,14 and 15, revealed coincidences in at least the initial part of the sequence of the events.

As a further control on whether these coincidences could be the result of the specific pair of participants or if they might have been obtained even with a different partner, we compared the data of each of the five "senders" with the data of all their nineteen unpaired "receivers".

For pair 13 we obtained similar or better results with "receivers" 7,8,10,14,16,17,18. For pair 8 we obtained identical results only with the "receiver" 10. For pairs 11,14 and 15 we did not obtained better or more precise coincidences. See results in Figure 2.

Pair 8:

Pair 11:

Pair 14:

Pair 15:
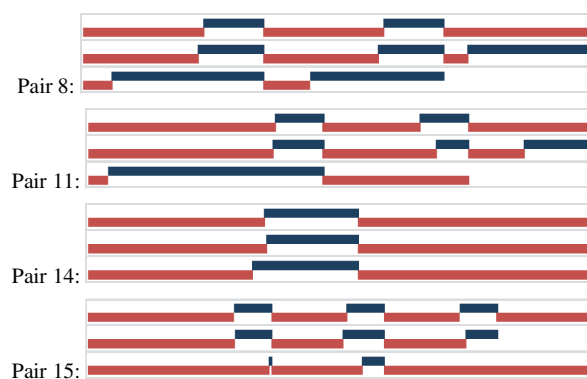
Figure 2. Results of the coincidences of silence (red color) and signal events (black color) detected by the BrainScanner™ between the pairs of participants. First row: the sequence of events of the protocol of stimulation; second row: sequence of events detected in the EEG activity of the "sender"; third row: sequence of events detected in the EEG activity of the "receiver" by using the classification model obtained analyzing the EEG activity of the "sender".

EEG power spectrum correlations

The Pearson's *r* correlation values with corresponding 95% CIs between the silence and signal events of each of the twenty pairs of participants separately for the five frequency bands, are reported in the Table S1 (see Supplementary Material). The corresponding graphs are available at http://figshare.com/articles/BBI_Confirmatory/1030617.

In Figure 3, we report the alpha band normalized power spectrum values recorded in the fourteen channels of the EEG activity of pair 15 as an example of strong correlation.

The overall average correlations among the twenty pairs transformed by using the *Fisher* r-to-z transformation formula, was estimated with 5000 bootstrap resamplings with the corresponding confidence intervals for each EEG frequency band, separately for the silence and signal events. The results are reported in Table 2.

As observed in the pilot study, we found reliable correlations in the alpha band for both silence and signal events and in the gamma band only for the silence events. In the pilot study we also observed the strongest correlation in the alpha band.

Fourteen out of the twenty pairs of participants showed statistically significant correlations in at least one of these two frequency bands.

Specificity analysis control

To control whether these correlations were specific to the pair of participants or might have also been obtained with their unpaired partners, we redid the 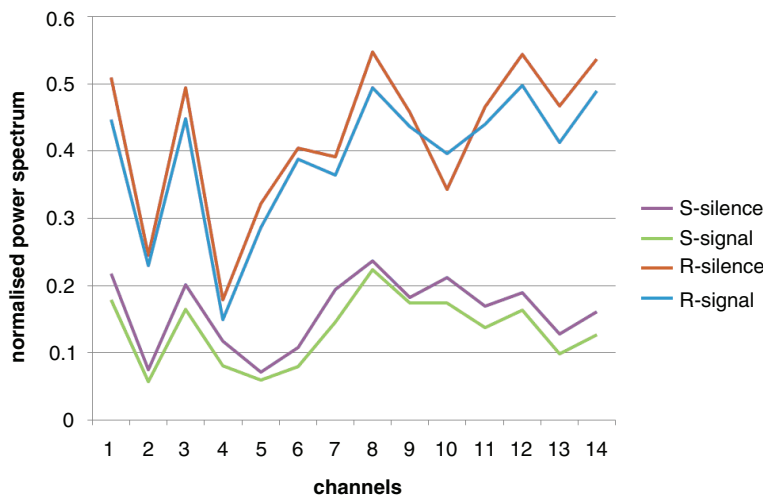correlation related to the Alpha and Gamma bands pairing each "sender" with all nineteen unpaired "receivers". The results were the follow: Alpha silence: 0.35; 95% CI: 0.25,0.43; Alpha signal: 0.27; 95% CI: 0.17,0.36; Gamma silence: 0.35; 95% CI: 0.25,0.43; Gamma signal: 0.14; 95% CI: 0.04,0.23.

General discussion

Compared with the pilot study of Tressoldi *et al.*, (2014), in the present study the pairs of participants were approximately 190 km away each other, the length of the sequence of events was randomized and the durations of the silence and signal periods were halved. However, the percentage of the overall correct sequences of events were almost identical with those observed in the pilot study. In the pilot study, the overall percentage of correct identification of the events was 78%; 95% CI=72–87 with respect to the 78.4%; 95% CI=68.7–85.7, observed in the present study.

Furthermore the average correlation estimated with 5000 bootstrap resamplings among all pairs of data was 0.58; 95% CI=0.46–0.69 and 0.55; 95% CI=0.43–0.65 for the alpha band respectively for the silence and the signal periods in the pilot study and 0.32; 95% CI=0.18–0.44, for silence and 0.27; 95% CI=0.13–0.40, for signal events in this confirmatory study. For the gamma band, the correlation values were 0.36; 95% CI=0.24–0.49 and 0.32; 95% CI=0.19–0.46 for the silence and signal, respectively, in the pilot study and 0.23; 95% CI=0.10–0.37 and 0.12; 95% CI=-0.009–0.26 in the present study.

The differences in the strength of correlations between the pilot and the present study may well be explained by the reduction of fifty percent in the duration of the silence and signal events with a consequent increment of the signal/noise ratio.

Figure 3. Normalized power values of the alpha band recorded on the fourteen channels of EEG activity, related to the silence and signal events for pair 15. Legend: S = Sender; R = Receiver.

Table 2. Averaged correlations with the corresponding confidence intervals for each EEG frequency band, separately for the silence and signal events. Statistically significant correlations are colored in bold.

	Delta		Theta		Alpha		Beta		Gamma	
	silence	signal	silence	signal	silence	signal	silence	signal	silence	signal
Correl	0.05	0.12	0.11	0.05	**0.37**	**0.31**	-0.05	0.02	**0.24**	0.14
95% CI	-0.04–0.14	-0.005–0.23	-0.01–0.23	-0.03–0.15	**0.21–0.54**	**0.13–0.40**	-0.16–0.05	-0.09–0.16	**0.10–0.37**	-0.01–0.26

The alpha band is a marker of attention (Klimesh, 2012; Klimesch *et al.*, 1998), whereas the gamma band is a marker of mental control as typically observed during meditation (Cahn *et al.*, 2010; Lutz *et al.*, 2004) and in this case the correlations we have observed could represent an EEG correlate of the synchronized attention between the pairs of participants.

We think that these results are mainly due to the innovative classification algorithm devised for this line of investigation and the enrolment of participants selected for their long friendship and experience in maintaining a mental concentration on the task. The drop of coincidences after three segments, corresponding to approximately five minutes, could be a limit of our classification algorithm to detect the differences between silence and signal, because of an increase of exogenous and endogenous EEG noise correlated to fatigue and loss of concentration (mental connection) between the two partners.

Effects not related to mental interaction at distance or artifacts that could explain our results

External effects
The large distance between the pair of participants excludes any sensorial connections between them. The only possibilities of artificial connections between the EEG activity of the pairs of participants could be caused by sensorial triggers sent to the participant with the role of "receiver" by the computer recording his/her EEG activity. This possibility was excluded because the randomization, both of the start of the delivery of the protocol and of the length of sequences of events, was controlled only by the computer connected with the EEG activity of the participant with the role of "transmitter" and no acoustic or visual events were associated with these computations. Another possible source of artifacts could derive from the research assistants managing the computers connected with the EEG activity of the two participants. In this case the only possibility of synchronizing the EEG of the two participants could be obtained if the research assistant who randomized the type of the sequence of events sent this information to the research assistant of the "receiver" who sent auditory signals to influence the EEG activity of the "receiver". All our research assistants were part of the research team and this possibility can be excluded with certainty.

Internal effects
Another potential cause of the observed correlations between the stimulation protocol and the EEG activity of the participants with the role of "receivers", could be due to their capacity to self-induce a synchronization of their EEG activity with the timing of the protocol delivered to the "sender" partner, predicting when it started, after 1, 2 or 3 minutes and when a silence or signal period was delivered. Apart from the fact that even if our participants were able to self-induce a differential EEG activity, they could guess the correct timing of the stimulation protocol only for one third of the sessions, and there is no evidence that humans can obtain such as skills for time sequences lasting 60 or 30 seconds. Furthermore, the participants that were also co-authors of this study, specified that they did not engage in such activity.

Statistical artifacts
Could our results be simply artifacts on how we analyzed the data?

The specificity control of the observed correlations between the Alpha and Gamma bands of the pairs of participants, casts doubt on their specificity. In other words, they could also be observed correlating the data of unpaired participants.

The change of the method to detect the coincidences, that is, by using the classification model obtained with the EEG data of the "senders" to measure the coincidences in the sequence of silence and signal events in the EEG of the "receivers" plus the specificity control, reveals potential true coincidences in only four out of the twenty pairs of participants.

Are these results sufficient to support the hypothesis that human minds and their brains, can be connected at distance? Only multiple independent replications and further controls on further potential methodological and statistical artifacts can support this hypothesis both using our data and different participants.

While awaiting new and independent controls and replications of our findings, we are planning to improve the current stimulation protocol to support a simple mental telecommunication code at distance. For example, it is sufficient to associate any small sequence of events with a message, i.e. silence-signal = "CALL ME"; silence-signal-silence = "DANGER", etc.

The next steps of this line of research are an optimization of the classification algorithm to detect longer sequences of events and the analysis of data online.

Data availability
figshare: BBI_Confirmatory, doi: http://dx.doi.org/10.6084/m9.figshare.1030617 (Tressoldi, 2014a).

Software availability
The BrainScanner™ classification software used in this study is available on request from Pasquale Fedele, email: p.fedele@liquidweb.it.

The ad-hoc software written in C++ for Windows 7 used to control the delivery of the choice of protocols and the timing of the EEG activity recordings is available under a CCBY license from *figshare*: Mind Sync Data Acquisition Software, doi: http://dx.doi.org/10.6084/m9.figshare.1108110 (Tressoldi, 2014b).

Supplementary Material

General description of the Supported vector machines (SVMs)

Supported vector machines (SVMs) are an example of generalized linear classifiers also defined as maximum margin classifiers because they minimize the empirical error of classification maximizing the margins of separation of the categories. SVMs can be considered as alternative techniques for the learning of polynomial classifiers very different to the classical techniques of neural networks training.

Neural networks with a single layer have an efficient learning algorithm, but they are useful only in the case of linearly separable data. Conversely, multilayer neural networks can represent non-linear functions, but they are difficult to train because of the number of dimensions of the space of weights, and because the most common techniques, such as back-propagation, allow to obtain the network weights by solving an optimization problem not convex and not bound, consequently it presents an indeterminate number of local minima (Basheer & Hajmeer, 2000). The SVM training technique solves both problems: it is an efficient algorithm and is able to represent complex non-linear functions. The characteristic parameters of the network are obtained by solving a convex quadratic programming problem with equality constraints or box type (in which the value of the parameter must be maintained within a range), which provides a single global minimum.

Additional references

The results of a comprehensive search of all studies related to this line of research revealed at least eighteen studies from 1974 until the present time. These references are presented in a Word document as part of the Supplementary Material.

Figure S1. A participant wearing the complete apparatus used during the experiment: Emotiv™ EEG, digital glasses and headphones.

Table S1. Correlation values and 95% confidence intervals between the EEG activity of each set of paired participants for each EEG frequency band and the two classes of events, silence and signal. Values in bold are statistically significant (when the confidence intervals do not include the zero).

Pair	Delta		Theta		Alpha		Beta		Gamma	
	silence	signal	silence	signal	silence	signal	silence	signal	silence	signal
1	0.04	0.29	0.25	**0.51**	0.1	0.42	0.03	-0.02	0.48	0.31
	-0.56,0.82	-0.29,0.71	-0.12,0.63	**0.07,0.78**	-0.69,0.84	-0.37,0.83	-0.55,0.72	-0.54,0.49	-0.07,0.87	-0.29,0.73
2	-0.18	-0.40	0.26	-0.14	0.25	0.01	-0.32	-0.46	-0.21	-0.02
	-0.78,0.41	-0.83,0.52	-0.14,0.67	-0.47,0.31	-0.53,0.74	-0.57,0.66	-0.65,0.20	-0.76,0.64	-0.51,0.20	-0.51,0.83
3	0.09	-0.22	0.25	-0.02	0.23	-0.23	-0.07	-0.04	0.2	0.25
	-0.44,0.42	-0.61,0.39	-0.47,0.61	-0.41,0.39	-0.07,0.65	-0.66,0.26	-0.34,0.41	-0.54,0.65	-0.19,0.61	-0.31,0.72
4	-0.14	**-0.57**	**0.5**	-0.15	0.07	-0.30	0.3	0.03	**0.48**	0.26
	-0.70,0.48	**-0.86,-0.01**	**0.15,0.84**	-0.36,0.65	-0.69,0.70	-0.81,0.35	-0.35,0.70	-0.62,0.45	**0.04,0.81**	-0.24,0.67
5	0.007	0.64	0.01	0.33	0.04	0.53	0.14	0.47	**0.49**	0.46
	-0.65,0.63	-0.18,0.95	-0.43,0.70	-0.19,0.65	-0.43,0.58	-0.17,0.90	-0.35,0.60	-0.58,0.84	**0.00,0.77**	-0.11,0.85
6	-0.15	0.001	0.1	0.05	0.17	0.26	-0.14	-0.03	0.07	-0.16
	-0.72,0.57	-0.77,0.72	-0.36,0.68	-0.37,0.49	-0.32,0.72	-0.36,0.78	-0.66,0.42	-0.74,0.58	-0.47,0.66	-0.61,0.64
7	0.11	0.13	0.07	-0.23	0.45	0.13	0.28	0.21	**0.6**	**0.58**
	-0.36,0.56	-0.35,0.79	-0.57,0.64	-0.72,0.53	-0.11,0.75	-0.30,0.62	-0.34,0.70	-0.31,0.67	**0.03,0.83**	**0.16,0.82**
8	-0.32	-0.05	0.28	0.15	**-0.48**	-0.25	0.06	0.1	0.38	0.18
	-0.84,0.20	-0.56,0.55	-0.37,0.72	-0.45,0.61	**-0.89,-0.04**	-0.71,0.42	-0.41,0.48	-0.48,0.75	-0.31,0.80	-0.47,0.75
9	0.38	0.14	-0.10	-0.15	**0.45**	0.13	0.37	0.23	**0.65**	**0.6**
	-0.12,0.79	-0.49,0.65	-0.45,0.44	-0.48.0.34	**.012,0.85**	-0.35,-0.64	-0.18,0.74	-0.35,0.68	**0.32,0.86**	**0.04,0.85**
10	0.34	0.33	-0.23	-0.07	**0.24**	0.16	-0.29	-0.26	0.27	0.32
	-0.45,0.75	-0.44,0.72	-0.62,0.50	-0.52,0.50	**.004,0.80**	-0.02,-0.69	-0.75.0.25	-0.69.0.15	-0.24,0.66	-0.23,0.72
11	0.3	-0.08	-0.42	-0.05	**0.84**	**0.63**	0.06	-0.05	0.51	0.21
	-0.09,0.73	-0.59,0.58	-0.81,0.11	-0.67,0.46	**0.66,0.93**	**0.19,0.89**	-0.65.0.66	-0.50,0.36	-0.56,0.95	-0.50,0.57
12	-0.06	-0.006	-0.03	-0.13	**0.44**	**0.38**	-0.06	-0.10	-0.09	-0.25
	-0.59,0.59	-0.57,0.68	-0.22,0.42	-0.43,0.38	**0.14,0.85**	**0.11,0.85**	-0.63,0.75	-0.62,0.79	-0.38,0.30	-0.69,0.47
13	0	0.36	-0.09	-0.11	0.22	**0.68**	-0.10	-0.13	-0.23	-0.27
	-0.67,0.58	-0.15,0.82	-0.56,0.39	-0.59,0.30	-0.40,0.70	**0.33,0.89**	-0.57,0.68	-0.65,0.67	-0.51,0.43	-0.52,0.01
14	0.25	0.29	0.06	0.28	**0.53**	0.16	0.04	**0.7**	0.26	0.45
	-0.25,0.62	-0.18,0.74	-0.55,0.65	-0.23,0.66	**0.16,0.82**	-0.22,0.62	-0.35,0.43	**0.15,0.89**	-0.42,0.72	-0.24,0.86
15	0.18	0.32	0.14	0.06	**0.64**	**0.73**	-0.25	-0.11	**0.25**	-0.002
	-0.56,0.66	-0.38,0.75	-0.50,0.65	-0.39,0.49	**0.18,0.87**	**0.46,0.88**	-0.58,0.31	-0.64,0.68	**0.007,0.71**	-0.40,0.81
16	-0.09	0.26	0.11	0.26	**0.38**	0.27	-0.36	0.08	**-0.35**	**-0.37**
	-0.58,0.37	-0.26,0.71	-0.24,0.56	-0.09,0.68	**0.05,0.74**	-0.23,0.68	-0.79,0.07	-0.44,0.58	**-0.64,-0.01**	**-0.65,-0.06**
17	0.03	0.39	-0.46	-0.15	**0.69**	**0.65**	**-0.53**	**-0.57**	0.04	**-0.52**
	-0.49,0.45	-0.01,0.70	-0.81,0.17	-0.67,0.31	**0.52,0.84**	**0.31,0.87**	**-0.83,-0.11**	**-0.85,-0.14**	-0.62,0.50	**-0.74,-0.27**
18	0.29	0.1	0.52	0.41	**0.62**	**0.55**	0.006	0.06	0.5	0.07
	-0.30,0.77	-0.39,0.63	-0.007,0.94	-0.09,0.78	**0.21,0.86**	**0.18,0.85**	-0.40,-0.44	-0.39,0.64	-0.20,0.82	-0.52,0.63
19	**0.31**	**0.34**	**0.64**	-0.10	**0.59**	**0.52**	0.28	0.51	0.68	0.54
	0.006,0.74	**0.07,0.77**	**0.34,0.83**	-0.55,0.31	**0.26,0.82**	**0.11,0.87**	-0.51,-0.71	-0.32,0.87	-0.19,0.91	-0.35,0.89
20	-0.35	0.21	0.38	0.4	-0.02	0.05	**-0.49**	-0.03	-0.20	-0.09
	-0.72,0.02	-0.27,0.70	-0.21,0.82	-0.21,0.79	-0.37,0.39	-0.51,0.58	**-0.76,-0.15**	-0.44,0.34	-0.68,0.50	-0.51,0.49

References

Acunzo DJ, Evrard R, Rabeyron T: **Anomalous experiences, psi and functional neuroimaging.** *Front Hum Neurosc.* 2013; **7**: 893.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Basheer IA, Hajmeer M: **Artificial neural networks: fundamentals, computing, design, and application.** *J Microbiol Methods.* 2000; **43**(1): 3–31.
PubMed Abstract | **Publisher Full Text**

Bishara AJ, Hittner JB: **Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches.** *Psychol Methods.* 2012; **17**(3): 399–417.
PubMed Abstract | **Publisher Full Text**

Cahn BR, Delorme A, Polich J: **Occipital gamma activation during Vipassana meditation.** *Cogn Process.* 2010; **11**(1): 39–56.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Chang CC, Lin CJ: **LIBSVM: a library for support vector machines.** *ACM TIST.* 2011; **2**(3): 27.
Publisher Full Text

Cox TJ: **Scraping sounds and disgusting noises.** *Appl Acoust.* 2008; **69**(12): 1195–1204.
Publisher Full Text

Cumming G: **The new statistics why and how.** *Psychol Sci.* 2014; **25**(1): 7–29.
PubMed Abstract | **Publisher Full Text**

Filk T, Römer H: **Generalized quantum theory: Overview and latest developments.** *Axiomathes.* 2011; **21**(2): 211–220.
Publisher Full Text

Grau C, Ginhoux R, Riera A, *et al.*: **Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies.** *PLoS One.* 2014; **9**(8): e105225.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Klimesch W: **α-band oscillations, attention, and controlled access to stored information.** *Trends Cogn Sci.* 2012; **16**(12): 606–617.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Klimesch W, Doppelmayr M, Russegger H, *et al.*: **Induced alpha band power changes in the human EEG and attention.** *Neurosci Lett.* 1998; **244**(2): 73–76.
PubMed Abstract | **Publisher Full Text**

Lutz A, Greischar LL, Rawlings NB, *et al.*: **Long-term meditators self-induce high-amplitude gamma synchrony during mental practice.** *Proc Natl Acad Sci U S A.* 2004; **101**(46): 16369–16373.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nosek BA: Open Science Collaboration. **An open, large-scale, collaborative effort to estimate the reproducibility of psychological science.** *Perspect Psychol Sci.* 2012; **7**(6): 657–660.
Publisher Full Text

Pais-Vieira M, Lebedev M, Kunicki C, *et al.*: **A brain-to-brain interface for real-time sharing of sensorimotor information.** *Sci Rep.* 2013; **3**: 1319.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rao R, Stocco A: **Direct Brain-to-Brain Communication in Humans: A Pilot Study**. 2013.
Reference Source

Steinwart I, Christmann A: **Support vector machines**. Springer, 2008.
Publisher Full Text

Tressoldi P: **BBI_Confirmatory.** *figshare.* 2014a.
Data Source

Tressoldi P: **Mind Sync Data Acquisition Software.** *figshare.* 2014b.
Data Source

Tressoldi PE, Pederzoli L, Caini P, *et al.*: **Brain-to-brain (mind-to-mind) interaction at distance: a pilot study**. 2014.
Publisher Full Text

Von Lucadou W, Romer H: **Synchronistic phenomena as entanglement correlations in Generalized Quantum Theory.** *J Consc Stud.* 2007; **14**(4): 50–74.
Reference Source

Wagenmakers EJ, Wetzels R, Borsboom D, *et al.*: **Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011).** *J Pers Soc Psychol.* 2011; **100**(3): 426–432.
PubMed Abstract | **Publisher Full Text**

Wagenmakers EJ, Wetzels R, Borsboom D, *et al.*: **An agenda for purely confirmatory research.** *Perspect Psychol Sci.* 2012; **7**(6): 632–638.
Publisher Full Text

Walach H, von Stillfried N: **Generalised quantum theory—basic idea and general intuition: a background story and overview.** *Axiomathes.* 2011; **21**(2), 185–209.
Publisher Full Text

Open Peer Review

Current Referee Status:

Sam Schwarzkopf
Institute of Cognitive Neuroscience, University College London, London, UK

I thank the authors for this newly revised version of their manuscript. Unfortunately, while they carried out some of the analyses we discussed, the evidence still does not support their claims, even though the claims have now been toned down. However, before I elaborate on this point I feel I must first discuss a more general point.

It is not a secret that I am an outspoken critic of 'psi' research, that is, experiments seeking to demonstrate the existence of purported extra sensory perception, telepathy, psychokinesis, precognition etc. My prior belief in such phenomena is presumably very different from that of the authors of this manuscript. But this raises an important question: What level of evidence should suffice to accept their findings and approve this manuscript?

The way I see it, the purpose of this review is not to evaluate the claim that there is a telepathic link (the authors don't use this term but this is what is implied) between brains at any distance. The authors would need far more compelling evidence to convince me that this is actually a possibility but this is not really the claim put forth by this manuscript. Rather the claim is that there is an as yet *unexplained phenomenon that appears like a telepathic link.* The authors have however been very clear from the beginning that future research needs to strengthen this evidence. They write in the conclusions: 'Only multiple independent replications and further controls on further potential methodological and statistical artifacts can support this hypothesis…'

But of course this statement applies to *any* scientific result. Only replication and appropriate controls can really strengthen our confidence in a hypothesis. However, this does not imply that we should accept every small result as tentative or preliminary result. Even an infinite number of replications would be completely worthless if the basic premise of the experiment is not sound. My review is therefore not about the evidence for the telepathy hypothesis but about the claim that something as-yet unexplained has been observed at all. I do not think the evidence supports this statement for the following reasons:

1. Main analysis is meaningless

Most of my previous reviews focused on a major conceptual flaw in the main analysis used by the authors. They decode the presence of a signal using data within each individual and randomly splitting the data samples into training and test sets. My reanalysis with two different non-linear classifiers (a SVM similar to that used by the authors as well as a k-nearest-neighbour algorithm) suggests that training and testing on completely arbitrary stimulus labels produces almost perfect decoding performance, provided

that the epochs are not too short (a few seconds seem to suffice). Presumably this is because temporal correlations between adjacent time samples (which I previously demonstrated exist in the raw data) will break down when labels are defined completely at random.

The authors have not addressed this issue at all. They do not report any attempt to repeat my analysis using their classifier but only state that they repeat their original analysis five times to show it is robust. This does not address the issue, however, because it does not matter how the training and test data sets are randomised. It will produce similar findings every time because of temporal correlations between adjacent samples.

In essence, this means that the main analysis, which remains the primary result in the new version of the manuscript, is completely confounded. Unless the authors can provide evidence that this does not happen for their own classifier, which to my understanding is based on a publicly available implementation of SVM (LibSVM) that in my experience does not differ dramatically from the one I used in my reanalyses, and, if so, a compelling explanation why that might be the case, I do not see any way this analysis can be accepted. It simply shows that decoding within participant works due to temporal correlations. Therefore this entire analysis should be removed.

2. Cross-participant analysis is confounded

The authors performed an analysis we discussed in the comments on the previous version of this manuscript. They used data from the Sender in each pair to train the classifier and then decode using the data recorded from the Receiver. This would be a better test because it shouldn't be confounded by temporal correlations in each participant's time series and it also directly tests the question whether there is actually any similarity in the EEG signals between the two brains. In my previous attempts to perform such an analysis, I only did this on a handful of pairs and found no evidence of above chance decoding. This is consistent with the author's new results that this Sender-to-Receiver decoding only works for 5 of the pairs. As I said in my previous review, considering that there are some correlations between the time series for different participants it isn't terribly surprising that there should also be some cross-participant decoding.

The authors therefore also performed a control analysis by training on Senders from these 5 'successful' pairs showing cross-participant decoding but testing on unpaired Receivers. They report that training on Sender 13 results in similar accuracy when testing the classifier on data from Receiver 7, 8, 10, 14, 16, 17, and 18. What they do not mention is that these are all Receivers for whom the first stimulus period was perfectly correlated with that of Sender 13. For Sender 8 they report only good correspondence with Receiver 10 and for the remaining three 'successful' pairs they reported no correspondences. This in itself appears questionable because inspection of the NewCoincidences files suggest that there appear to be some rather obvious 'coincidences' in the decoding for training on Sender 8 with Receivers 2, 4, 5, 7, 9, 14, 17, 18, and 20. There are also many traces for which the decoded neither classified the trace as signal nor as silence. It still is unclear to me what this means. Typically a SVM classifier will assign either one or the other label so it is puzzling why these data are missing. By and large this also again raises the question why the authors chose to classify events based on these 'coincidences' instead of actually quantifying the classification performance directly.

Regardless of these issues, what these data show is that cross-participant decoding is not very impressive. Even if we accept that cross-participant decoding works only for 4 of the 20 pairs this still does not really suggest that something very unusual is happening. This only actually reflects a small number of events being classified correctly – by my count 5 out of 34 signal events, if we follow the

authors way to quantify 'coincidences'.

Importantly, the authors also perform a similar cross-pair analysis to quantify the strength of correlations between time series from different participants. This clearly suggests that the correlations they observed for the alpha and gamma bands *within* pairs are actually extremely close (and with overlapping confidence intervals) as those observed *between* pairs. This suggests that there are simply strong correlations in the EEG time series between different people regardless of whether they were paired up or not. These correlations are particularly strong for the alpha band, which makes sense considering the relaxing experimental conditions under which the recordings were conducted. It may also reflect anticipatory behaviour or imagery the participants could have engaged in – a notion the authors reject as impossible although they didn't actually test this possibility. Or perhaps it could also relate to slow fluctuations in gamma power that this experimental design didn't control for because the stimulus protocol was always 30 s of signal followed by 60 s of silence. In any case, considering these correlations across pairs it seems not overly surprising that there should be some decoding when training on Senders and testing on Receivers in some pairs.

3. Other outstanding issues

The authors also made no effort to address some of the perhaps more minor concerns. The binomial test on the decoding when both signal and silence events are combined is still statistically incorrect. It simply is not true that the chance level is 50% because there were fewer signal events than silence events. For unbalanced designs a classifier will frequently not assign events/trials with a probability of 50%. The appropriate way to test for significant decoding would be to use a permutation test. The fact that the Bayes Factor in this test was 360,625 should raise suspicion. Rather than being interpreted as overwhelming evidence such a result is usually either completely trivial or a hint of an artifact.

The authors also still claim that the initial silence period was randomised to be 1, 2, or 3 minutes, which according to the traces is clearly not correct. The initial silence was either 1 or 2 minutes which means there was less randomisation than the authors claim. The authors also do not even begin to discuss the high colinearity between the stimulus protocols used in the 20 pairs and how this could have influenced the predictability of the design or even the correlation between whether any given epoch was a signal or silence period and the state of the EEG time series. The authors previously suggested that it would apparently be impossible for participants to predict the sequence. I am not convinced by this argument but even if we take it at face value this does not preclude order effects.

Conclusion

Even if some of these issues could be addressed by a further revision, as I discussed here and in my previous reviews, the numerous issues with inadequate randomisation and lack of control conditions lead me to think that only a complete overhaul of this experiment could begin to address these problems properly. More importantly, as this is now the third revision of the manuscript I feel confident in saying that the discussion is going in circles (two points are insufficient evidence to confirm circularity but three are just about enough). I am happy I was given the chance to review this manuscript. Looking into these raw data has been educational for me and this has been my first experience with a totally transparent review process. I have long advocated making review comments public.

However, I believe a reviewer's job is to provide an expert's opinion about manuscript. It is not to make a judgment about whether a study should or should not have been published. With the *F1000Research* publishing model this is the decision of the readers, and I also do not believe that such a decision should

be made solely based on one reviewer's comments. In a situation like this, where two reviewers clearly disagree, ideally a third expert opinion should be sought to adjudicate. Alternatively, the readers may decide to weight the two reviews based on what they perceived is fair and accurate.

In any case, I do not think I can contribute any more useful information to this discussion. I believe I have evaluated this manuscript as thoroughly as I can. Perhaps I missed or misunderstood something fundamental. The beauty of the transparent review process is that anyone can identify these shortcomings and use them to reach their final decision.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 24 Oct 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

On behalf of all authors, I agree with you that this review process must be considered very important, useful, but it is now time to end it. All potential interested readers can read our comments, see and use the raw data and results to achieve an independent opinion.

I would end by thanking you for all your efforts to write detailed and constructive comments that will help us to improve our methodological and statistical approaches in the future experiments.

Competing Interests: I'm the corresponding author

Version 2

Referee Report 30 September 2014

doi:10.5256/f1000research.5711.r6254

❌ **Sam Schwarzkopf**
Institute of Cognitive Neuroscience, University College London, London, UK

The authors have clarified several of the questions I raised in my previous review. Unfortunately, most of the major problems have not been addressed by this revision. As I stated in my previous review, I deem it unlikely that all those issues can be solved merely by a few added paragraphs. Instead there are still some fundamental concerns with the experimental design and, most critically, with the analysis. This means the strong conclusions put forward by this manuscript are not warranted and I cannot approve the manuscript in this form.

1. The greatest concern is that when I followed the description of the methods in the previous version it was possible to decode, with almost perfect accuracy, any arbitrary stimulus labels I chose. See http://dx.doi.org/10.6084/m9.figshare.1167456 for examples of this reanalysis. Regardless of whether we pretend that the actual stimulus appeared at a later time or was continuously alternating between signal and silence, the decoding is always close to perfect. This is an

indication that the decoding has **nothing to do with the actual stimulus** heard by the Sender but is opportunistically exploiting some other features in the data. The control analysis the authors performed, reversing the stimulus labels, cannot address this problem because it suffers from the exact same problem. Essentially, what the classifier is presumably using is the time that has passed since the recording started.

2. The reason for this is presumably that the authors used non-independent data for training and testing. Assuming I understand correctly (see point 3), random sampling one half of data samples from an EEG trace are **not independent data**. Repeating the analysis five times – the control analysis the authors performed – is not an adequate way to address this concern. Randomly selecting samples from a time series containing slow changes (such as the slow wave activity that presumably dominates these recordings under these circumstances) will inevitably contain strong temporal correlations. See TemporalCorrelations.jpg in http://dx.doi.org/10.6084/m9.figshare.1185723 for 2D density histograms and a correlation matrix demonstrating this.

3. While the revised methods section provides more detail now, it still is unclear about exactly what data were used. Conventional classification analysis report what data features (usual columns in the data matrix) and what observations (usual rows) were used. Anything could be a feature but typically this might be the different EEG channels or fMRI voxels etc. Observations are usually time points. Here I assume the authors transformed the raw samples into a different space using principal component analysis. It is not stated if the dimensionality was reduced using the eigenvalues. Either way, I assume the data samples (collected at 128 Hz) were then used as observations and the EEG channels transformed by PCA were used as features. The stimulus labels were assigned as ON or OFF for each sample. A set of 50% of samples (and labels) was then selected at random for training, and the rest was used for testing. Is this correct?

4. A powerful non-linear classifier can capitalise on such correlations to discriminate arbitrary labels. In my own analyses I used both an SVM with RBF as well as a k-nearest neighbour classifier, both of which produce excellent decoding of arbitrary stimulus labels (see point 1). Interestingly, linear classifiers or less powerful SVM kernels fare much worse – a clear indication that the classifier learns about the complex non-linear pattern of temporal correlations that can describe the stimulus label. This is further corroborated by the fact that when using stimulus labels that are chosen completely at random (i.e. with high temporal frequency) decoding does not work.

5. The authors have mostly clarified how the correlation analysis was performed. It is still left unclear, however, how the correlations for individual pairs were averaged. Was Fisher's z-transformation used, or were the data pooled across pairs? More importantly, it is not entirely surprising that under the experimental conditions there will be some correlation between the EEG signals for different participants, especially in low frequency bands. Again, this further supports the suspicion that the classification utilizes slow frequency signals that are unrelated to the stimulus and the experimental hypothesis. In fact, a quick spot check seems to confirm this suspicion: correlating the time series separately for each channel from the Receiver in pair 1 with those from the Receiver in pair 18 reveals 131 significant ($p<0.05$, Bonferroni corrected) out of 196 (14x14 channels) correlations… One could perhaps argue that this is not surprising because both these pairs had been exposed to identical stimulus protocols: one minute of initial silence and only one signal period (see point 6). However, it certainly argues strongly against the notion that the decoding is any way related to the mental connection between the particular Sender and Receiver in a given pair because it clearly works between Receivers in different pairs! However, to further control for this possibility I repeated

the same analysis but now comparing the Receiver from pair 1 to the Receiver from pair 15. This pair was exposed to a different stimulus paradigm (2 minutes of initial silence and a longer paradigm with three signal periods). I only used the initial 3 minutes for the correlation analysis. Therefore, both recordings would have been exposed to only one signal period but at different times (at 1 min and 2 min for pair 1 and 15, respectively). Even though the stimulus protocol was completely different the time courses for all the channels are highly correlated and 137 out of 196 correlations are significant. Considering that I used the **raw data** for this analysis it should not surprise anyone that extracting power from different frequency bands in short time windows will also reveal significant correlations. Crucially, it demonstrates that correlations between Sender and Receiver are artifactual and trivial.

6. The authors argue in their response and the revision that predictive strategies were unlikely. After having performed these additional analyses I am inclined to agree. The excellent decoding almost certainly has nothing to do with expectation or imagery effects and it is irrelevant whether participants could guess the temporal design of the experiment. Rather, the results are almost entirely an artefact of the analysis. However, this does not mean that predictability is not an issue. The figure StimulusTimecourses.jpg in http://dx.doi.org/10.6084/m9.figshare.1185723 plots the stimulus time courses for all 20 pairs as can be extracted from the newly uploaded data. This confirms what I wrote in my previous review, in fact, with the corrected data sets the problem with predictability is even greater. Out of the 20 pairs, 13 started with 1 min of initial silence. The remaining 7 had 2 minutes of initial silence. Most of the stimulus paradigms are therefore perfectly aligned and thus highly correlated. This also proves incorrect the statement that initial silence periods were 1, 2, or 3 minutes. No pair had 3 min of initial silence. It would therefore have been very easy for any given Receiver to correctly guess the protocol. It should be clear that this is far from optimal for testing such an unorthodox hypothesis. Any future experiments should employ more randomization to decrease predictability. Even if this wasn't the underlying cause of the present results, this is simply not great experimental design.

7. The authors now acknowledge in their response that all the participants were authors. They say that this is also acknowledged in the methods section, but I did not see any statement about that in the revised manuscript. As before, I also find it highly questionable to include **only** authors in an experiment of this kind. It is not sufficient to claim that Receivers weren't guessing their stimulus protocol. While I am giving the authors (and thus the participants) the benefit of the doubt that they actually believe they weren't guessing/predicting the stimulus protocols, this does not rule out that they did. It may in fact be possible to make such predictions subconsciously (Now, if you ask me, **this** is an interesting scientific question someone should do an experiment on!). The fact that all the participants were presumably intimately familiar with the protocol may help that. Any future experiments should take steps to prevent this.

8. I do not follow the explanation for the binomial test the authors used. Based on the excessive Bayes Factor of 390,625 it is clear that the authors assumed a chance level of 50% on their binomial test. Because the design is not balanced, this is not correct.

9. In general, the Bayes Factor and the extremely high decoding accuracy should have given the authors reason to start. Considering the unusual hypothesis did the authors not at any point wonder if these results aren't just far too good to be true? Decoding mental states from brain activity is typically extremely noisy and hardly affords accuracies at the level seen here. Extremely accurate decoding and Bayes Factors in the hundreds of thousands should be a tell-tale sign to check that there isn't an analytical flaw that makes the result entirely trivial. I believe this is what

happened here and thus I think this experiment serves as a very good demonstration for the pitfalls of applying such analysis without sanity checks. In order to make claims like this, the experimental design must contain control conditions that can rule out these problems. Presumably, recordings without any Sender, and maybe even when the "Receiver" is aware of this fact, should produce very similar results.

Based on all these factors, it is impossible for me to approve this manuscript. I should however state that it is laudable that the authors chose to make all the raw data of their experiment publicly available. Without this it would have impossible for me to carry out the additional analyses, and thus the most fundamental problem in the analysis would have remained unknown. I respect the authors' patience and professionalism in dealing with what I can only assume is a rather harsh review experience. I am honoured by the request for an adversarial collaboration. I do not rule out such efforts at some point in the future. However, for all of the reasons outlined in this and my previous review, I do not think the time is right for this experiment to proceed to this stage. Fundamental analytical flaws and weaknesses in the design should be ruled out first. An adversarial collaboration only really makes sense to me for paradigms were we can be confident that mundane or trivial factors have been excluded.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.

Referee Response 30 Sep 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

Regarding pt 7 it was pointed out to me that the authors state in the methods 'Seven healthy adults, five males and two females, were selected for this experiment and included as co-author.'

I apologise that I somehow missed this acknowledgement in the manuscript when I read it for my review. However, this doesn't change any of the conclusions about this point. Familiarity with the paradigm is a problem here. The fact that the participants were involved in the study is exacerbating this, but it would even be a problem if they were all outsiders because each person participated repeatedly and often in consecutive sessions.

Moreover, one might question the temporal order of events in this statement. Were participants truly selected first and then offered authorship, or were they authors/research assistants to begin with who all agreed to participate? The latter seems a lot more likely also considering that the authors were the same in the pilot study.

Either way, this is a side point. The most critical problem, which I already pointed out for the previous version, is that the results are almost certainly a trivial analytical artifact and that even if there were some link between brains these methods are wholly inadequate to detect it.

Competing Interests: No competing interests were disclosed.

Author Response 30 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Adversial collaboration proposal

"An adversarial collaboration only really makes sense to me for paradigms were we can be confident that mundane or trivial factors have been excluded."

My proposal in precisely in this sense. We are ready to change our paradigm. Why do you not agree on an experimental design which can exclude most "mundane or trivial factors" to pre-register on the OSF?

Competing Interests: I am the corresponding author

Referee Response 01 Oct 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

"Why do you not agree on an experimental design which can exclude most "mundane or trivial factors" to pre-register on the OSF?"

For three reasons:

1. Time. I think engaging in peer review is an important activity, it is particularly important to scrutinise controversial and/or extraordinary claims like this, and I am also learning new things from this experience. Yet I am actually also an active research scientist and I should really do the research I am funded to do! So if I invest much of my time outside of my main research activities it should be reasonably limited and also have high probability of providing interesting results, which brings me to...

2. Conceptual issues. From what I can see, the experimental hypothesis as it stands at the moment does not seem to make much sense to me in and of itself. Surely if you want to decode a signal transmitted from one brain to another, you should have a design that directly tests for the presence of such a connection. So rather than decoding the labels of a stimulus protocol delivered to the Sender in the brain activity of the Receiver, isn't what you really want to use the Sender's brain activity to predict the brain activity in the Receiver? This, if appropriately controlled, could theoretically provide evidence for a sychronisation in EEG signals due to the "mental link".

3. Technical issues. Considering the issues with the present data I am also not entirely sure how feasible that whole idea even is. Seeing that raw signal traces are apparently correlated between different people regardless of whether they were paired up as Sender or Receiver, which stimulus protocol was used, and at what time the recording took place, I can't currently conceive of a good way to test your hypothesis. Even if you estimate the baseline level of "mundane" association it is probably difficult to distinguish that from associations due to experimental factors. Before we consider improvements of this design I think the source of this trivial association first needs to be understood properly.

To be honest, this whole process has been very educational for me (and hopefully for others). I am not aware of anyone having used this random sampling classification method with EEG data before but if so this should serve as a deafening wake-up call that this procedure is fundamentally flawed. There has been a lot of discussion of the problem of non-independence in classification analysis in the fMRI literature (in fact, a new study about this problem was just published by Russ Poldrack's lab). In my mind, your experiments demonstrate how important these considerations are and generally how critical it is to have experimental designs with appropriate baseline/control

conditions. My suggestion would be to completely revise the current manuscript to use this as a hands-on illustration of these issues. Collecting additional data from control conditions could round this off nicely - but I don't think you need an adversarial collaboration for that.

That said, I am very happy to have a discussion about a possible experimental design that could actually try to test this hypothesis and that when this is finalised it could then be pre-registered at the OSF. However, I don't think this should be just me. Certainly, such a discussion would benefit from being completely public allowing other researchers, including those who are more familiar with EEG than me, to contribute.

A proper adversarial collaboration (such as the one between Richard Wiseman and Marilyn Schlitz) to me implies that both parties are directly involved in carrying out the experiments in order to ensure that everything is done exactly in accordance with the protocol. Such a collaboration is a significant investment in terms of time and effort and so I think a very clear idea of the experiment and what answers it can possibly provide therefore must come first. I am not opposed to that in principle, but considering my own time constraints I would presently only consider this for more established designs like Bem's precognition tests or the presentiment experiments.

Competing Interests: No competing interests were disclosed.

Author Response 02 Oct 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Proposal:
to rule out that the results of our classifier can be due to the non-independence of data for training and testing, what do you think if we train the classifier with the data of the Sender and test its predictive accuracy on the data of the Receiver and vice versa?

Competing Interests: I'm the corresponding author

Referee Response 02 Oct 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

"to rule out that the results of our classifier can be due to the non-independence of data for training and testing, what do you think if we train the classifier with the data of the Sender and test its predictive accuracy on the data of the Receiver and vice versa?"

I think this would certainly be a good start. Based on my conversations with colleagues this is in fact what most people (including myself) believed you had done! It only became clear to me that you didn't when I started looking at the data - the 50% statement also gave it away of course.

However, as I did mention in the review comments to version 1 when I tried doing this, that didn't work so I expect this will be a null result (try it though please - I didn't use your classifier code after all).

On the other hand, due to the correlation between the time series of even unpaired participants it may work under certain conditions - but that would again simply indicate that some variance in the signal is common between people simply due to the general experimental conditions. So

essentially you need to conduct a permutation test in which you train on one sender and test on all the receivers in turn to see what the baseline level of decoding accuracy is (the same approach would also be sensible for the correlation analyses). For that to work the correlations between the stimulus protocols must be abolished of course - another reason to have more randomisation in the stimulus protocols.

Competing Interests: No competing interests were disclosed.

Referee Response 02 Oct 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

...the 50% statement also gave it away of course.

Sorry this sounds misleading. I am not insinuating that this was fraudulent or that you were trying to hide something. I just mean that from the methods it wasn't clear how the analysis was done so most people assumed that you used Sender's EEG to decode the Receiver's EEG. While you have added some detail now, it still could be clearer on that point.

Competing Interests: No competing interests were disclosed.

Author Response 02 Oct 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Agreed.

I will train the classifier with the data of each sender at a time and I'll test the prediction accuracy on all the receivers' data. The null hypothesis is that the prediction accuracy will be almost identical independently from the receiver role (paired or not) and the length of the protocol.

I will do a similar permutation test by using the power spectrum data of only the alpha and gamma bands.

I'll post the results and version 3 of the paper ASAP.

Competing Interests: I'm the corresponding author

Referee Response 02 Oct 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

I think the first test could just be whether decoding works at all between different brains (i.e. Sender->Receiver). As I mentioned, this wasn't working for me so if that doesn't work, the permutation analysis is presumably irrelevant.

Competing Interests: No competing interests were disclosed.

Author Response 13 Oct 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

On behalf of all co-authors, I've posted the new version of the paper with the results of the agreed procedure to control potential statistical artifacts in the method used for the detection of coincidences and the controls for the specificity of the results.

As you read, the results obtained with the new procedure will be less impressive than those obtained with the original procedure, but sufficient for us, who see the "glass half-full", to continue this line of research.

Competing Interests: I'm the corresponding author

Author Response 21 Oct 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Sorry I did not mention where I added the new analyses in the paper.

On page 6, there is the new paragraph "Control of potential statistical artifacts" related to the new procedure to classify the coincidences of events between the pairs of participants:

On page 7, there is the new paragraph "Specificity analysis control" related to the control of the specificity of the correlations related to the alpha and gamma bands.

On page 8, the paragraph "Statistical artifacts", has been expanded.

Competing Interests: I'm the corresponding author

Referee Response 23 Oct 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

Thank you for the response. I will try to work on this within the next few days. First a quick question though: I assume the raw data files for the experiment won't have changed since the previous version (or the first one, except for where the erroneous files were replaced)?

Another issue is the classifier. Would you kindly send me the actually classifier function? For the limited re-analyses I conducted previously I used a different classifier (SVM implemented in Matlab Bioinformatics toolbox as well as a custom-made k-nearest-neighbour algorithm). I would like to compare this to the actual method you used to understand where any discrepancies come from.

Competing Interests: No competing interests were disclosed.

Author Response 12 Nov 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Sorry for replying late to to this comment since I had not activated the "Tracking" option.
The raw data are the same of v.2 after the correction of the erroneous files.
As to the our BrainScanner classifier, it is based on the libraries available here

Version 1

Referee Report 09 September 2014

doi:10.5256/f1000research.4642.r6065

✗ **Sam Schwarzkopf**

Institute of Cognitive Neuroscience, University College London, London, UK

This study aims to test the rather unusual hypothesis that the brains of two individuals separated geographically by almost 200 km can form a telepathic link that is measurable with EEG. While this is arguably an implausible hypothesis, it is certainly testable. Unfortunately, I believe this is not an adequate experiment to test such a hypothesis. There are major problems both with the experimental design and the statistical analysis. While the authors may be able to address some of my comments with additional analyses and control experiments, the sheer number of issues demands a complete overhaul of the entire study. I broke up my concerns into three sections:

1. Non-naive participants and predictability of the protocol:

1. There were only 7 participants in this experiment. This means that in order to collect the 20 data sets (ignoring 2 excluded ones) every person participated in multiple recording sessions, both as "sender" and as "receiver" (except for one person (subject A) who acted only once as sender but thrice as receiver). Therefore even if participants were naive about the experimental protocol before their first recording session, in subsequent sessions they would be very familiar with both the arousing auditory stimulus (sound of a crying baby) and with the sequence of events (30s stimulus/signal periods interspersed by 60s periods of silence). In fact, the "senders" were even explicitly told about this sequence at the start of each session. Moreover, typically the roles of "sender" and "receiver" were reversed for a pair of participants right after their first recording. Thus the knowledge of the design would have been fresh in the second "receiver's" mind.

2. On this note, based on the subject initials in the data spreadsheets I wonder if the first author participated in this experiment? This is not explicitly stated so perhaps it is simply a coincidence that the initials match. All of the other subject initials also match those of other authors, however, so at the very least it should be acknowledged who (if any) among the participants were authors. Of course, it can be entirely acceptable to take part in your own study but this should always be reported in the methods and it very much depends on the experimental design. Certainly, for a study of this kind, where the predictability of experimental events is critical, I would be very concerned how an in-depth knowledge of the experimental protocol affects the results. It would certainly make the claim somewhat questionable that participants could not have known the randomisation of the protocols.

3. Perhaps the biggest improvement over the pilot experiment, in which the sequence and duration of the stimulus protocol was always fixed and thus **completely predictable**, is the fact that the overall duration of the protocol was randomised (from three options, i.e. protocols with 1, 2, or 3 stimulus periods) and that the duration of the initial silence period was apparently randomised between 1-3 minutes (but see point 1.4). Thus the 'receiver' should have been less able to predict the exact onset of the first stimulation period. However, after this initial onset the protocol was always fixed (i.e. 30s stimulus periods separated by 60s silence), so provided they could make a reasonable guess about the onset of the first stimulus period, the rest of the session would still have been rather predictable.

4. Inspection of the traces of the stimulus protocols suggests that for the three different protocols the sequence of stimulation events was always perfectly aligned with the onsets of several other protocols. This is because the randomization of the initial silence period was not somewhere between 1-3 minutes as implied in the methods (actually this states "seconds", but the corresponding author already acknowledged that this is a typo). Instead, as far as I can tell the initial silence period appears to have been **either** 1 min or 2 min. This of course means that the onset of the first stimulus period was either been at 1 min (12/20 sessions) or 2 min (8/20 sessions). As discussed in point 1.3, after this onset the sequence would have been fairly predictable to the participant. The duration of the initial silence was therefore the most unpredictable part of the experiment for the "receiver" – provided they had no other cues (see point 1.5) or had prior knowledge of the randomisation (see points 1.2 and 1.6).

5. Is it conceivable that there were any cues for the "receiver" whether this session had 1 min or 2min of initial silence? It is unclear from the description of the methods whether the picture of the "sender" appeared in the goggles from the beginning of the recording session, including the initial silence period (but I assume this was the case). What could the participants hear and feel from the experimental room, noises made by the attending research assistants, etc?

6. Assuming that the timestamps on the spreadsheets indicate the timing of recording, it can be seen that for two thirds of the pairs, the duration of the initial silence period in the second recording was the opposite of that in the first recording. Since these pairs were just reversing the role of "sender" and "receiver", the "receiver" could then be predisposed to expect a shorter/longer initial silence period in the second recording compared to the first. So even if "receivers" always assumed that the onset of the stimulus was the opposite of the first session they would have been correct more than half of the time (see also my discussion of incorrect statistical assumptions in point 3). Moreover, blocks of sessions typically had one participant in common (sessions 1-4 subject F, sessions 5-10 subject D, sessions 11-13 subject A, 14-30 subject PT). It thus seems likely that "receivers" were implicitly aware of the randomisation of the onsets.

7. Regarding the predictability of the sequence, is it possible that there were any time cues helping the "receiver" keep the timing and thus predict the sequence of stimulus events? While the participants were wearing headphones and goggles, could they have heard the ticking of a clock or a dripping tap or other regular noises (perhaps from the experimental equipment)? Could there have been signals in other sensory modalities (floor vibrations, air flow in the room)? Such cues need not even be external for the participants could have kept time, e.g. by using their respiration. In particular considering that all participants had experience with meditation, yoga, or similar practices this does not seem unrealistic.

8. As discussed, the only aspect of the experiment that was comparably unpredictable (except for the potential caveats discussed in the previous points) was the onset of the initial silence period. Subsequently, the sequencing of stimulus and silence was fixed and it was actually fairly unimportant whether the overall protocol duration was short (1 stimulus), medium (2 stimuli), or long (3 stimuli) because participants would only have to maintain the fixed rhythm of 30s stimulus followed by 60s silence. The fact that decoding becomes progressively worse for segments later in the protocol (as shown by Tables 1a-c) may thus be a result of the "receiver's" inability to maintain the rhythm as the session progressed. This is in part supported by some of the traces in which the classifier detected stimulus periods that considerably exceeded 30s in the latter half of the session (in particular, session "tLrPT"). The deterioration of decoding accuracy could also be due to the uncertainty as to whether there would be more stimulus periods or not because the participant could not be sure that they were in a short, medium, or long session.

In summary, there were multiple problems with familiarity of participants with the experimental paradigm and the predictability with the rhythm of stimulus and silence periods. To address this, the experiment should have been made much more unpredictable with properly randomized onsets and jittered durations for all the silence events.

2. The nature of the decoded signal:

1. The participants' familiarity with the crying baby stimulus also raises further questions. First, it somehow undermines the whole idea of transmitting information between two brains. Except for the first session all participants already knew that the stimulus was a crying baby. This makes transmitting that information redundant. More importantly, it also means that participants could have been imagining (or at least thinking about) the crying baby sound at regular intervals prescribed by the experiment. In that case, the classifier algorithm would have simply decoded the thoughts/imagery or the mental effort of the "receiver" to receive the crying baby sound. Combined with the issues with predictability of the sequence discussed in point 1, this would make the results hardly surprising. One way to address this would be to have two very distinct signal events and training the classifier to distinguish those in addition to the silence period (e.g. a crying baby vs a calming surf). I can however understand that the authors focused only on binary events (stimulus or silence) but in that case at the very least they would have to address the concerns with the predictability of the stimulus sequence discussed in point 1.

2. A related problem with the decoding analysis is that there is no way of knowing whether the decoded signal has anything to do with the "sender's" experience of a crying baby. Was there any debriefing of participants? Did any of the "receivers" hear a crying baby during the recording session? Or perhaps that is expecting too much. Did they at the very least report the feeling of receiving any information from the sender? One way to control for this would have been to have sessions both with and without a "sender" (obviously randomised so that the "receiver" could not know) and to see if the classifier still identifies stimulus and silence periods at these regular intervals.

3. The previous suggestion would also help to address another concern about the nature of the decoded signal. As the authors themselves (briefly) acknowledge in the discussion, the alpha and gamma frequency bands are markers of attentional engagement, arousal, or mind wandering. Thus the decoding might instead simply exploit the temporal evolution of the EEG signal over the course of the session related to these factors. While this does not entirely explain why the decoding is so high for the first stimulus period (but see discussion of this problem in point 1), it

certainly would be an alternative explanation for why decoding becomes progressively worse over the course of the session (see also point 3.5).

3. Incorrect statistical assumptions and questions about analysis:

1. The statistical analysis used for testing whether decoding performance was above chance levels is incorrect, because the authors did not take into account that this is an unbalanced design. Therefore, contrary to the authors' description in the methods the expected chance level is not 50%. Because stimulus periods made up less of the overall duration than silence periods, even if the classifier consistently (and incorrectly) assigned the silence label decoding accuracy would be greater than 70% (i.e. the proportion of silence within a session). The propensity of the classifier to choose one over another class label is also not necessarily 50%, especially in unbalanced designs. The use of a standard binomial test against 50% chance performance is therefore not correct. Instead the authors should have used a permutation test that estimates the true chance performance under these conditions.

2. The "coincidences" measure used by the authors is also questionable. It is not immediately clear whether the definition of overlapping segments could have inflated the decoding results somehow. It seems odd that this measure was used at all considering that it should be straightforward to compare the traces directly. It also seems strange that there was a subjective disagreement between the two raters – the definition of coincidences sounds pretty simple.

3. The methods do not provide nearly sufficient detail to understand how the decoding analysis was performed. The authors state that they used PCA to reduce the dimensionality of the EEG channels but they do not state what data were actually used for classification. Was it the band-pass filtered EEG signal trace within short time windows? Was it the frequency-power spectrum within each time window? How long were the time windows? Or was a sliding time window used?

4. In this context it is also quite odd that the classifier performs so consistently. It certainly seems very odd that the classifier would hardly ever misclassify the initial silence period or that there are never any gaps within stimulus periods. Physiological data are typically very noisy. It is surprising to see such reliable classification even for the "sender", let alone the "receiver". Again, this is difficult to understand without a clear idea of how exactly the classification was performed.

5. It is also unclear what the classifier was trained on. The authors state that a randomly selected "fifty percent of these data" were used for training. Did the authors use 50% of the data for training separately for each participant in the pair and then test the classifier on the remaining 50%? This would be incorrect because there are likely to be temporal correlations between adjacent data points (again, this would be a lot clearer if we knew what data were actually used). Or were these 50% from the "sender" and then used to classifier 100% of data from the "receiver"? It seems more defensible to assume that the "sender" and "receiver" are statistically independent (unless of course the hypothesis of a telepathic link is true). However, the temporal proximity of data points might still be a concern even in this case (see also points 2.3). Especially considering the fact that the authors used one of the most powerful linear SVM kernels (radial basis function) for classification, it is very unclear what attributes in the data the classifier exploited. One of the main problems with such multivariate decoding analyses is in fact that it is entirely opportunistic – the algorithm will find the most diagnostic information about the class labels in the data to produce the an accurate classification without any regard to whether this diagnostic information is actually meaningful to the hypothesis. So without any better understanding of what was done, it is quite

plausible that the classification simply decoded how much time had passed since the start of the experiment. One way to reveal this would be to rerun the classification with different class labels that are orthogonal to the stimulation sequence. If the classifier exploited some attribute about the temporal evolution of the signal it should still perform well under those circumstances.

6. The lack of methodological detail also makes it impossible to understand the description of the correlation analyses. It is stated that recordings were broken up into 4s time bins. How does this relate to the correlations that were calculated? In Figure 2 alpha power in different channels are plotted. This does however not indicate which of the periods (i.e. first, second or third stimulus, or average across them?) these power values came from (at least according to my count pair 15 should have had three stimulus periods). It also does not explain whether this is just the data from one of the 4s bins or an entire segment or if it was averaged across all segments?

7. How were the correlations listed in Table 2 averaged across pairs? Was it taken into account that the same participants contributed to several of these correlations? Moreover, was any correction for multiple comparisons applied to the number of frequency bands?

8. Several of the decoding traces for the "receiver" contain zeros. Two examples are in fact shown in Figure 1 in the top and middle rows. What does that mean? Was the recording simply stopped at that point? If so, why? The stimulus protocol with the sender was still running at that time?

9. The authors state that the recording for the "receiver" was triggered manually by the research assistant after receiving the signal via the internet from the lab with the "sender". Would this not introduce an uncontrolled lag in the recordings? Surely it should be technically feasible to automate this and trigger the recording simultaneously (or at least with a fixed lag due to the internet transmission)?

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 11 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Many thanks indeed for your accurate review and the time spent on our paper. The reply to all your and the second reviewer comments and a new version of the paper wil be posted ASAP.

Competing Interests: I'm the corresponding author

Referee Response 11 Sep 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

After writing this review a number of additional issues have come to my attention. For the review I had merely inspected the decoding traces included in the spreadsheets in the Coincidences folder. However, inspection of the actual raw data files reveals the following problems:

Decoding arbitrary stimulus labels

1. We loaded a few of the raw data and applied decoding analysis using a non-linear SVM classifier with RBF kernel. As described in my review, the methods provide insufficient detail to understand what was done. Nevertheless, it is easy to replicate almost perfect decoding (98-99% correct) of the stimulus label using what seems to be the procedure described in the manuscript. Critically, however, this works for arbitrarily chosen stimulus labels that are completely independent from the actual stimulation sequence (this is the control analysis I suggested to the authors in point 3.5 of my review). Examples of such decoding analysis can be seen at: http://dx.doi.org/10.6084/m9.figshare.1167456

This demonstrates that the decoding has nothing to do with the stimulation sequence or with any purported connection between the brains of Sender and Receiver – rather the classifier exploits temporal correlations in the recorded signal because training and test data are not truly independent. Therefore even if the authors controlled for all of the other concerns I raised in the review, the general analysis approach is simply incorrect for addressing this question.

Irregularities in the raw data files

2. The time stamps for Sender and Receiver do not match for Pair 3 (4 hours time difference) and Pair 17 (70 minutes time difference).

3. In fact, for Pair 3 the file for the Receiver (subject PT) and the file for the Sender in Pair 16 are **exactly identical** (except for the stimulus condition labels in column 3). We can speculate that this is due to a copying error and that the Receiver in Pair 3 should have been subject F.

4. For Pair 12 the Sender is listed as subject P but it presumably should have been subject PT - according to the Coincidences files subject P was never paired as Sender with subject A as Receiver.

Non-naive participants

5. The subject names in the raw data files essentially confirm that **all** participants in this experiment were also authors on this manuscript (see point 1.2 in my review). Again, it can be acceptable to include some authors in studies and in some situations it may even be defensible if all participants are authors, especially if the design is truly unpredictable and the experiment studies a very basic process. However, this should be acknowledged **explicitly** in the methods, and it seems difficult to justify this in an experiment such as this where the participants should be naive with regard to the experimental protocol.

Competing Interests: No competing interests were disclosed.

Author Response 12 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

May you clarify which events you classified with your non-linear SVM classifier? Segments corresponding to the stimulation protocols (i.e. silence 60 sec; signal 30 sec.) or different ones?

Competing Interests: I am the corresponding author

Referee Response 15 Sep 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

Thank you for your question. I concede that the decoding analysis I performed is very rudimentary. I merely took the raw data observations, that is, columns 4-17 (counting from left) in the raw data spreadsheets are the input data. I understand these are the raw signals from the 14 channels. As labels I either used the stimulus labels in the third column or arbitrarily assigned labels that are uncorrelated with the true stimulus. Following the description in your methods section I then randomly sampled 50% of the observations (rows) in the data and the labels to be used as training set and the other 50% as testing set.

This approach works well for decoding the presence or absence of the "stimulus" regardless of whether you train and test on the same data set (that is, sender or receiver) or whether using 50% of the sender as training set and the *other* 50% of the receiver as testing set (the latter shows modestly reduced decoding accuracy). Presumably it should work also when using data from different pairs in particular for those who had identical stimulus protocols. I haven't tried that yet though.

It does depend on the length of the "stimulus" periods you choose. Using a random stimulus that assigns the stimulus as being on or off for each sample (row) in the data does not work. It does therefore depend on having slow stimulus periods presumably because the classifier exploits slow temporal correlations.

The main difference I can see between mine and your analysis (in so far I understood it correctly) is that you first transformed the 14 data columns into eigenvalues using PCA. But the lack of this step cannot explain the artifactual decoding in my reanalysis. PCA merely reduces the dimensionality of the data set by taking into account the covariance between the 14 channels and expressing the data in this lower dimensional space (although you didn't specify how many principal components you used in the decoding). This may in fact explain why the decoding periods are slightly less noisy in your case - although I'm still surprised by that.

Or did you use another set of observations, say, short time windows rather than the individual samples as I did here? If so this should be described in more detail as it is far from clear in the present form. Even so, however, any data transformation you may have used would I think preserve the underlying problem: decoding works for arbitrary stimuli. This is why it also isn't surprising that the presumably erroneous data set in Pair 3 can be decoded well even though the wrong subject (sender data from Pair 16) seems to have been used here with the incorrect stimulus labels. The same applies to the pairs which were recorded with large time delays, that is Pairs 17 and 3 (although this time delay may simply be due to the incorrect data file for the receiver).

Unlike the first reviewer I think it is entirely premature to speculate about "brain-to-brain" links to operate across time - rather I would say that recordings at different times (different days) would serve as excellent control data sets. My prediction, according to what we have seen so far, is that decoding using this method will work very well even across different days and across different subject pairs because it simply relies on the temporal evolution of the slow wave EEG signal under these experimental conditions.

To sum up all of my (admittedly very extensive) criticisms, I think the one major overarching problem with this experiment is that it lacks an appropriate control. There is no experimental

condition that allows you to pit the telepathy hypothesis against alternative explanations. There simply is no way of knowing whether the decoding has anything to do with the sender, with the task, with the stimulus, etc. Unless I made an enormous oversight, the reanalysis I attempted strongly suggests that decoding has nothing whatsoever to do with the stimulus protocol. Even if decoding works across Sender and Receiver (or even across participants from different pairs) there is no way to rule out that it simply exploits the temporal evolution of the signal, which is presumably similar under these conditions.

I know this is harsh, but given these problems I don't see any other way to address these criticisms than to completely redo the experiment using a properly controlled design.

Competing Interests: No competing interests were disclosed.

Referee Response 15 Sep 2014

D. Sam Schwarzkopf, UCL Experimental Psychology & Institute of Cognitive Neuroscience, UK

Correction: In my previous comment I said that the decoding works even when training on Sender and testing on the Receiver. This is however not correct. I looked at that analysis briefly but I didn't plot back the decoding. Of course, the classification was >80% correct but that is simply because the classifier in this case incorrectly classifies everything as silence. Chance accuracy in this case is of course exactly at that level. This is in fact the error with incorrectly assuming a chance level of 50% I mentioned in my review and it shows you how easy it is to misinterpret such results!

Therefore, the most critical thing you need to clarify is 1) the nature of the data you used to classify, 2) what exactly was chosen as training and test observations, and 3) was this done separately for Sender and Receiver?

Competing Interests: No competing interests were disclosed.

Author Response 18 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

1. Non-naive participants and predictability of the protocol concerns.
Reply: You correctly pointed out the participants were not naïve to the experimental procedure, but this was an intentional selection and in the second version of the paper we disclosed they were all co-authors. The choice of these participants was justified in the Method section. Only future studies will confirm if the same results can be observed including no-selected participants.

The matter that they could self-generate a differential EEG activity corresponding to the timing of three protocols of stimulation was your major concern.
We think it could be possible if all the following conditions are satisfied:

- They trained a lot to self-induce a differential EEG modulation activity following the timing of our protocol: one type of EEG activity for 1 min, another type for 30 sec, repeatedly;

- They were able to synchronize the above EEG activity with the timing of the protocol delivered to the "sender" partner, predicting when it started, after 1,2 or 3 minutes.

As to point a) we assure that none of the participants achieved this ability and applied it during the experiment. Furthermore it is the opinion of Dr. Giovanni Mento, an expert in the EEG correlates of

time estimation, is that at present there is no evidence a human adult could obtain such ability for time segments above few seconds, in particular because the precision of time estimation worsens with the length of the times to estimate, following the Weber law (e.g. Piras & Coull, 2011). As to point b) the randomization of the stimulation protocol start should predict a correct synchronization for only 33% of the trials.

Instead, as far as I can tell the initial silence period appears to have been either 1 min or 2 min.
Reply: In the "Stimulation protocol" paragraph, we clarified that the randomization of the onset of the stimulation protocol was among 1,2 or 3 minutes.

Is it conceivable that there were any cues for the "receiver" whether this session had 1 min or 2min of initial silence? ....**What could the participants hear and feel from the experimental room, noises made by the attending research assistants, etc?**
Reply: the "receiver" saw the image of the "sender" before the formal start of the stimulation protocol and remained visible until the end of the session. After the PC connected to the EEG was activated, the research assistant remained outside the room where the "receiver" was placed. The sound attenuated lab and the headphones filtered all external noises. No visual or auditory cues were available to be used to predict the sequence of events.

Since these pairs were just reversing the role of "sender" and "receiver", the "receiver" could then be predisposed to expect a shorter/longer initial silence period in the second recording compared to the first.
Reply: If participants engaged in an intentional guessing of the exact onset of the stimulation protocol, they could guess correctly 1/3 of the sessions. As clarified in point 1, our selected participants guaranteed the adherence of the instructions.

The nature of the decoded signal:
.....A related problem with the decoding analysis is that there is no way of knowing whether the decoded signal has anything to do with the "sender's" experience of a crying baby.

Reply: the initial project aimed at comparing implicit EEG activity with explicit reporting by the "receiver". However we did not find stimulus, both visual and auditory that could be presented for 30 sec without inducing an habituation apart the "baby crying" clip. However we aim at investigating the implicit/explicit reporting differences.

Incorrect statistical assumptions and questions about analysis:
....the expected chance level is not 50%.
Reply: In effect the definition of 50% as the expected chance level is misleading, given that we estimated only the percentage of coincidences overall and for silence and signal segments separately, and computed the Bayes Factor comparing the hypothesis that the percentage of coincidences were above the percentage of errors and missing.

The authors state that they used PCA to reduce the dimensionality of the EEG channels but they do not state what data were actually used for classification
Reply: in the version 2 we have clarified that we used the raw EEG data of participants with the role of "receivers" without any pre-processing or filter.

Did the authors use 50% of the data for training separately for each participant in the pair and then test the classifier on the remaining 50%? This would be incorrect because there are likely to be temporal correlations between adjacent data points.......it is quite plausible that the classification simply decoded how much time had passed since the start of the experiment. One way to reveal this would be to rerun the classification with different class labels that are orthogonal to the stimulation sequence. If the classifier exploited some attribute about the temporal evolution of the signal it should still perform well under those circumstances.

Reply: In the "Classification algorithm" paragraph we clarified that even if the EEG recording of both the sender and the receiver have been analyzed with the BrainScanner, only the results of the participants in the role of "receiver" were reported. We confirm we used 50% of the data for training the classifier together with the labels defining the silence and signal periods. In the pilot study, we observed that this percentage and the reduction of the signal periods to be analyzed, worsened the classification accuracy. The possibility that with 50% of the data used for the training it may be possible to predict the remaining 50% simply because of a temporal correlation between adjacent points is probabilistically possible if the randomization selected only or the majority of even or odd numbers. Repeating the classification five times we ruled out this possibility.

Reversing the classification labels yields the same results, as the task of the classifier is simply to detect differences between the two categories of events if any, but it is "blind" with respect to their meaning or nature.

..correlation analyses....**This does however not indicate which of the periods (i.e. first, second or third stimulus, or average across them?) these power values came from**

Reply: In the "correlation analysis" paragraph we specified that the normalized Power Spectral Density was calculated by collapsing all silence and signal periods.

How were the correlations listed in Table 2 averaged across pairs? ..... Moreover, was any correction for multiple comparisons applied to the number of frequency bands?

Reply: The overall correlation results presented in Table 2 are obtained by averaging the correlations presented in Table S1, to take in account individual (pair) differences. For the statistical approach we used, now disclosed in the paragraph "Statistical approach", familywise error rate controls don't apply.

Several of the decoding traces for the "receiver" contain zeros.....**What does that mean? Was the recording simply stopped at that point?**

Reply: In the examples presented in Figure 1 we clarified that these were considered as missing. As explained in the text, these are the outputs of our classifier that used all data of the participants with the role of "receivers" recorded applying the three stimulation protocols.

Irregularities in the raw data files
Reply: Many thanks for the accurate control of our raw data. Now we have uploaded the correct files.

I would say that recordings at different times (different days) would serve as excellent control data sets.
Reply: we agree that the next step is to devise a control condition where we expect no correlation and coincidences. Apart your suggestion, we could add "no stimulation" trials where no visual or

auditory information are delivered interspersed with "stimulation trials" like those used in the present study. Given we are devising a new pre-registered study, would you agree to have an "adversial collaboration" with us?

Competing Interests: I am the corresponding author

Referee Report 12 August 2014

James Lake
Arizona Center for Integrative Medicine, University of Arizona, Tuscon, AZ, USA

Thank you for the opportunity to review and comment on this important paper. This unique pilot study provides a strong beginning case for direct brain-to-brain communication. Further research along these lines incorporating advanced EEG analysis and brain-computer interface (BCI) technologies will lead to important insights about mechanisms underlying innate human Psi abilities while also developing practical approaches that will enhance the human capacity to both 'detect' and 'send' discrete units of 'information' at a distance.

The protocol used to create a maximally 'arousing' stimulus in the brain of the stimulated person is a novel approach to Psi research based on the (reasonable) assumption that the sound of a crying baby is very distressing to most humans and thus may represent a 'maximally arousing' stimulus. In a future study it would be interesting to measure EEG activity in a matched control group to determine whether or in what ways background EEG activity may change when the control group is exposed to the recorded sound of a baby crying. I believe it would add rigor to your methodology to develop a 'normative' data set of EEG responses to the stimulus (ie, through recordings from a representative cross-section of different ages and genders, etc) and to compare this 'normative' response to findings from the pilot study and future research findings.

I am interested to know how you decided on the stimulus duration and interval between stimuli. I am also interested to know how you determined the total duration (15 minutes). Were you concerned with subject fatigue? Did you arrive at these parameters from insights gained during the training phase? I am curious as to whether you might have obtained more robust correspondences using either a longer stimulus duration or a longer experiment duration. Your comments on this would be very useful. In terms of the algorithm of checking EEG data between 'sender' and 'recipient' for apparent above-chance correspondences, did you examine data during simultaneous recording epochs only, and did you consider examining the data for possible correspondences off-set in time? There appears to be an implicit assumption in your methodology that any significant correspondences between 2 (or more brains) will take place simultaneously or in the same general time 'period.' Along these lines it would be interesting to design statistical methods that could test for a time 'displacement' effect that may take place in your model of direct brain-to-brain communication.

I did not completely understand the goals of the 'training phase' or how this is done. It would be helpful to have a clearer and more detailed description of this part of the study. My understanding is that subjects are initially 'trained' to achieve robust responses to the 'baby cry' stimulus after which they enter into the experiment so that it will then be more likely that they will experience EEG responses of the type your algorithm is designed to detect. By extension, if the EEG response following stimulation by the 'baby cry' sound is robust and consistent across subjects, is your assumption that this stimulus would more likely

'evoke' a distant 'signal' in the receiver? In this case is 'training' really about trying to optimize/maximize signal to noise ratio in the 'distant signal' that may entangle two (or more) brains following the baby cry stimulus?

If I understand correctly, the classification algorithm built into the "BrainScanner" software comprises the critical aspect of your study in that the strength of your findings rests on the assumption that the correspondences in EEG activity obtained using the BrainScanner software and the Emotiv Neuroheadset are equivalent to real-time changes in EEG activity between two or more subjects monitored in parallel. Along these lines it would be helpful to clarify why the algorithm built into the "BrainScanner" software is the most appropriate algorithm for the study goals you have in mind. A more complete explanation of this point would also help me better understand your arguments for validating the significance of findings of apparent correspondences between the selected stimulus (ie, sound of baby crying) and general or specific changes in EEG activity in the 'receiver.'

Regarding the kind of information extracted from raw data, does the algorithm compare measures of discrete simple parameters like measures of amplitude, frequency and wavelength or more complex dynamic measures, for example cordance? Along these lines, if cordance is taken into account, does the algorithm analyze cordance with respect to mean differences between two or more brain regions? Does the algorithm average over different frequency or time domains, if so what are the parameters and threshold values and how are these derived and defended in developing your methods? Along these lines, it would be helpful in the discussion section to comment on the observation of 'apparent correspondences' between EEG traces in two or more humans because of common EEG dynamics, and in to include remarks that artifacts and shared EEG dynamics have been ruled out providing a compelling argument that findings cannot be 'explained away' by these phenomenon.

One potential concern along these lines is that the built-in software algorithm may be removing or simplifying particular EEG features prior to analysis which may in turn result in the *appearance* of greater uniformity and therefore the appearance of above-chance correspondences between two or more EEG tracings. In future studies it will be important to employ EEG apparatus and filtering software that directly address the possibility of apparent correspondences as an artifact of automated EEG filtering prior to analysis. The question concerns whether, in the absence of the software algorithm imposing 'structure' on more primitive signals, there would still be a statistically significant correspondence between two (or more) simultaneous recordings from two or more brains. Another phenomenon that that may give the appearance of significant real-time correspondences where none are present, and that should be directly addressed in future studies, has to do with the finding that a significant component of average EEG activity reflects widely shared EEG dynamics among the majority of humans in other words EEG activity between two or more humans probably varies in non-random ways. Such patterned non-random EEG activity may result in the appearance of correspondences when two or more brains are monitored. This issue has been raised in previous critiques of EEG (and fMRI) Psi studies. In order to build a compelling case that your findings do not incorporate bias and to verify a Psi effect you may first need to provide strong evidence ruling out artifacts or apparent correspondences between unique EEG traces due to widely shared EEG dynamics, as both may lead to the appearance of above-chance correspondences. This will require careful analysis of the recording method used, assumptions built into the software algorithm, and statistical methods used to test for significance.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 12 Aug 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Thank you for the accurate review and all comments to improve the present version. We will address all of them in the new version after we will receive the responses of the other referees.

For the moment we will reply to only some of the comments.

- *"...how you decided on the stimulus duration and interval between stimuli."*

  Before defining the present stimulation protocol, we ran multiple pilot trials. The goal was to identify a protocol simple to implement, short in duration (to avoid boredom and fatigue to the participant) and with sufficient signal to noise gain to be detected by our equipment. In the first formal pilot experiment, cited in the paper, we tested officially a first version of the protocol and the positive results prompted us to improve that protocol. On pag. 7 in the General Discussion we discuss these differences.

- *"I did not completely understand the goals of the 'training phase' or how this is done"*

  The training phase is not for the participants but only for the BrainScanner classifier. As for most classifiers, it is necessary to train the software to identify the best algorithm (both linear and nonlinear) to discriminate the two (in our case) class of EEG activity (silence and signal). After this phase, the software apply this algorithm to all data. We will clarify better this point in the next version.

- *"In order to build a compelling case that your findings do not incorporate bias and to verify a Psi effect you may first need to provide strong evidence ruling out artifacts or apparent correspondences between unique EEG"*

  In the new version we will expand this point. For the moment it is important to consider that the converging evidence supporting a real B-to-B interaction at distance is based on a replication of the results obtained in the pilot study even if using a shorter protocol of stimulation and the correlations in the alpha e gamma bands. If one looks at the graph of each single pair (available here), there are striking similarities in almost all pairs.

Competing Interests: I'm the corresponding author of the paper

Author Response 18 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

Did you examine data during simultaneous recording epochs only, and did you consider examining the data for possible correspondences off-set in time?

Reply: At present we do not have any plausible hypothesis about the possibility of a time-displacement effect, hence we are focusing our efforts to demonstrate a simultaneous correlation.

...it would be helpful to clarify why the algorithm built into the "BrainScanner" software is

the most appropriate algorithm for the study goals you have in mind....**does the algorithm compare measures of discrete simple parameters like measures of amplitude, frequency and wavelength or more complex dynamic measures, for example cordance?**

Reply: BrainScanner is one of the different classifiers available (see Amancio *et al.,* 2014) . These classifiers are one of the best options to classify signals in two or more categories. In our case for the input we used the results of the principal component analysis of the raw data, but it is possible to use other EEG characteristics as the input.

.... it would be helpful in the discussion section to comment on the observation of 'apparent correspondences' between EEG traces in two or more humans because of common EEG dynamics, and in to include remarks that artifacts and shared EEG dynamics have been ruled out providing a compelling argument that findings cannot be 'explained away' by these phenomenon......... **The question concerns whether, in the absence of the software algorithm imposing 'structure' on more primitive signals, there would still be a statistically significant correspondence between two (or more) simultaneous recordings from two or more brains.**........**you may first need to provide strong evidence ruling out artifacts or apparent correspondences between unique EEG traces due to widely shared EEG dynamics, as both may lead to the appearance of above-chance correspondences**

Reply: Some of these concerns were also shared by Dr Schwarzkopf. Our use of the raw data without any preprocessing or filtering and the fact that correspondences were observed only using those precise parameters (50% of random data from all silence and signal segments for training the classifier), and after five replications, offer a cautionary optimism about the reliability of our results. Moreover, in the Discussion we expanded the section related to possible artifacts. However if all agree that more experiments by us and independent authors are necessary both to confirm our findings and exclude alternative explanations of a nonlocal interaction.

Competing Interests: I am the corresponding author

Discuss this Article

Version 1

Author Response 18 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

I confirm that the coincidences are determined by comparing the sending protocol and the result of decoding the receiver's EEG.
In the v2 of the paper and in the replies to the 2 reviewers we have clarified this point.

Competing Interests: I'm the corresponding author

Reader Comment 17 Sep 2014

Gerard Ridgway, UCL Institute of Neurology, UK

Are the coincidences determined by comparing the sending protocol and the result of decoding the receiver's EEG, or by comparing the results from decoding sender's EEG and receiver's EEG? Figure 1 seems to suggest the former, but the text seems to suggest the latter. If the latter, do coincidences between false positives (i.e. wrongly predicted signal in sender and in receiver) and false negatives (wrongly predicted silence in both sender and receiver) count as coincidences?

Competing Interests: No competing interests were disclosed.

Author Response 08 Sep 2014

Patrizio Tressoldi, Dipartimento di Psicologia Generale, Università di Padova, Italy

"Should the statement that "the duration of the first silence segment was also randomized from one to three seconds" read "... one to three minutes"? Otherwise, I can't see that 2s of randomisation makes much difference to signal/silence blocks of 30s/60s."

Reply: You are right. This was a clear typo. We will correct "seconds" with "minutes" in rev.2

"In Figure 1, why do the durations of blocks appear different for different senders? I understood that the number of blocks was randomised (3, 5 or 7) and that the length of the first silence was randomly varied (as above), but that the durations of subsequent signal and silence blocks were fixed at 30s and 60s respectively. Are the different subjects' profiles not drawn to the same temporal scale?"

Reply: In the three examples presented in Figure 1, the first raws represents the stimulation protocol as delivered to the "sender". Apart the length of the first "silence" segment you are right, the should look identical. This was due to a formatting problem, but I confirm the silence segments lasted always 1 minute and the signal segments, 30 secs for all 20 sessions. The length of the segments in the second raws, may be different from the original protocol due to the imprecision of the classifier to identify the exact timing and duration of the silence and signal segments.

Thank you very much for your comments.

Competing Interests: I am the corresponding author

Reader Comment 05 Sep 2014

Gerard Ridgway, UCL Institute of Neurology, UK

Should the statement that "the duration of the first silence segment was also randomized from one to three seconds" read "... one to three minutes"? Otherwise, I can't see that 2s of randomisation makes much difference to signal/silence blocks of 30s/60s.

In Figure 1, why do the durations of blocks appear different for different senders? I understood that the

number of blocks was randomised (3, 5 or 7) and that the length of the first silence was randomly varied (as above), but that the durations of subsequent signal and silence blocks were fixed at 30s and 60s respectively. Are the different subjects' profiles not drawn to the same temporal scale?

Competing Interests: No competing interests were disclosed.