Two Attempted Retro-Priming Replications Show Theory-Relevant Anomalous Connectivity

JACOB JOLIJ

DICK BIERMAN

Department of Experimental Psychology, University of Groningen, The Netherlands

Abstract—We report a theory-relevant post hoc analysis of two Dutch retro-priming experiments that were part of a large replication project of the retro-priming experiment by Daryl Bem and colleagues. This replication project sought to investigate the role of the experimenter in psi studies. The results of the retro-priming experiments performed by student research groups at the University of Amsterdam (N = 61) and the University of Groningen (N = 222) did not, however, replicate Bem's earlier findings of an anomalous interference of a future stimulus on response times. We report the results of these two studies here, but the overall results will be reported elsewhere. Both Dutch studies used the exact same software as did Bem and colleagues. However, each study used a different questionnaire. The questionnaires asked for information that in previous research had been associated with success in psi tasks and that could help us to deal with individual differences, but above all could be used as selection criteria for participants in future studies. In the Amsterdam study, there were 14 questions, while in the Groningen study there were 55. A correlation analysis revealed several significant correlations between the psi effect in the Bem task and questionnaire items. In this paper we focus on the post-hoc research question: Is this global composition of the correlation matrix anomalous, as suggested by Generalized Quantum Theory? Rather than using the subjective number of 'significant' correlations as a dependent variable, we introduced two objective measures directly representing the correlation values in the cells to characterize the 'Connectivity' in the matrix. Our analysis revealed 'Connectivity' to be marginally significantly larger ($p < 0.075$) in the Groningen study and significantly larger in the Amsterdam study ($p < 0.025$). These results are discussed in the framework of the Consciousness Induced Restoration of Time Symmetry (CIRTS) theory and the Generalized Quantum Theory (GQT) that predict that as soon as there are reasons to expect replicability (or control) the extra chance psi scores will disappear or

reverse. The conclusion is that these results can be seen as support, though statistically not strong, for the GQT.

Introduction

In spite of strong experimental claims by some proponents, for instance Bem et al. (2015), of the psi hypothesis, the reality of claimed phenomena such as telepathy has not been accepted by mainstream science (e.g., Traxler et al. 2012). The major reason given for this skeptical position is that there appears to be no parapsychological experimental paradigm yet with a robust effect size that might be used by independent researchers to check this reality for themselves, i.e. given a specific power and a specific sample size, a specific distribution of outcomes as well as the probability for success in a single experiment can be calculated. Thus independent researchers can check if their experiments do follow the distribution of outcomes as predicted by the power of the original studies. This problem of nonreplicability is characteristic in controversial research areas, but during the last decade has also been a topic of intensive discussion in mainstream science because in a number of 'well-established' fields it turns out that the replicability rates are suspiciously low (Pashler & Wagenmakers 2012).

Within the field of parapsychology there are researchers who rather consistently seem to get results confirming the psi hypothesis, and there are independent researchers who fail to do so, and for these researchers improvement of the paradigm to a level where a specific power can be expected is high on the wish list.

The Heymansgroup of the University of Groningen, embedded in the Department of Experimental Psychology, developed a multi-year research program to that end (Heymansgroup.nl). If replication is such a problem, there has to be a large portion of unexplained and uncontrolled variance. This uncontrolled variance may be due to so-called questionable research practices or QRPs (Bierman, Spottiswoode, & Bijl 2016). However, as a result of further automation, especially the real time uploading of experimental data (Jolij & Bierman 2017), as now is becoming standard at the Heymansgroup and other research groups worldwide, it seems that these known QRPs are becoming increasingly unlikely.

The other obvious source for this error variance could be individual differences in participants. However, as is often claimed in experimental parapsychology, the uncontrolled variance may also be due to differences in experimenters. Therefore, the Heymansgroup embarked on the development of a selection instrument to be used in future experiments for participants *as well as experimenters* (see mission statement at Heymansgroup.nl). This selection instrument would then become publicly available in order to give

independent replicators the tools to reduce error variance due to individual differences in participants and experimenters.

To begin with, the group decided to use a coarse-grained approach to explore simple questions to help predict performance in a parapsychological experiment that was run in another context as sketched above.

Student experimenters had to first study the relevant literature searching for predictors of psi and come up with questions that captured the suspected predictor. The historical search for personality variables predictive of psi has had limited success. One of the successful and more consistent predictors of subjects' performance is 'belief in the reality of psi' (Schmeidler & McConnell 1958). However, even here the results seem ambiguous, often caused by nonbelievers scoring 'the wrong way'. Other predictors that have positively been implicated are 'experience with mental disciplines', 'creativity', 'intuitive versus rational thinking' (Honorton 1997).

Thus the questionnaires in both studies contained a number of questions relating to these potential predictors (see Appendix 1 for the wording of these questions). In spite of this study being highly explorative with regard to the questions intended to measure factors that had been associated with positive scoring in a psi task in the past, the student experimenters expected specific positive correlations in the literature they had been studying.

In the following we will describe the experiment in Groningen in detail in the Methods section. The experiment in Amsterdam that was done one year earlier was identical but used 14 different questions to be correlated with psi performance.

Methods

Background

The experiments are part of a large-scale, multiple-laboratory research effort by researchers unrelated to the University of Groningen or Amsterdam for which these Universities were invited to be participating research institutes. That particular large-scale experiment was trying to investigate the effects of the experimenter, and the results will be published elsewhere. The experiments were run by students as experimenters and they tried to replicate Bem's findings. On top of this evidential hypothesis, the students asked a separate research question of their own. Specifically: Can we find simple questions to be used in later research for subject selection for which the subject's responses would have a predictive value in the main psi task?

After the studies were completed, the supervisors in Groningen asked a further theory-driven question: Is the correlation matrix *as a whole* anomalous (is there explained variability than can be expected by chance)?

Bem's Retro-Priming Task

In this task (Bem 2011) adapted from de Boer and Bierman (2006), the normal sequence of prime and target in a 'valence judging' task is reversed. For instance, a picture of a flower is presented and the participant has to respond by a key press that the picture has a positive valence or another key press if the participant experiences a negative valence. Subsequently a 'prime' like the word 'ugly' (noncongruent condition) or 'beauty' (congruent condition) is presented. So the prime appears after a response has been given. In 'normal causal' priming the response times in the congruent condition are shorter. This can be explained by activation of the semantic network containing 'flower and beauty'. However, in the retro-priming experiment there is no causal model that would be able to account for response differences between the congruent and noncongruent conditions simply because the response is given earlier. The 'prime' can have an effect except when this effect is 'retrocausal'. The psi score then is operationalized as the difference between the mean response time in noncongruent trials and the mean response time in congruent trials. Details of the task such as the timing can be found in the original publication (Bem 2011).

Student Experimenters

Experimenters in Groningen were 11 psychology students (4 male and 7 female) ages from 21 to 23 who participated as part of their academic training. They had to subscribe to one out of a number of student projects and chose for this project one that was described with the following research question: *Can future events have an effect in the present?* (see Appendix 2 for full text). With regard to the educational goals, the supervisors emphasized the prevention of questionable research methods.

Task for the Student Experimenters

Each experimenter had to recruit his/her own participants, run the experiment using the standard software package as provided by Daryl Bem (rPrime, d.bem@gmail.com). They also participated in the construction of the questionnaire intended to measure aspects of the participant's personality that might predict results of the psi task. After data acquisition, they had to analyze their own results and analyze the global results with the whole group of experimenters. Finally, they had to write a scientific report.

Participants

Groningen participants were friends or family members of the student

experimenters. 143 were female with a mean age of 21.5 (SD = 1.6) while 79 were male with mean age 23.5 (SD = 4.6). They were compensated €5. In Amsterdam there were also more females (34; mean age = 25.8, SD = 12.0) than males (28; mean age = 28.3, SD = 12.9) and they were friends of the student experimenters who participated.

Tasks for the Participants

The participants first viewed a video clip, randomly selected from two such video clips. One clip was from a psi-proponent with an argument in favor of the reality of psi, and one clip was an argument from a well-known skeptical scientist against the reality of psi. This manipulation was embedded in the software and was of possible importance for the research into experimenter effects. The manipulation produced inconsistent results in Amsterdam and Groningen and will be evaluated over all participating labs by the principal investigators of the large-scale replication project.

After giving consent, the participants filled in the computerized questionnaire and subsequently ran the retro-priming experiment (see below).

Supervisors

The supervisor in Amsterdam was Eva Lobach, and the supervisors in Groningen were the authors of this article.

Materials

Questionnaire. The questions from the 5-point Likert scale (agree–don't agree) questionnaire were suggested by the student experimenters after they read the relevant literature about personality measures and performance in a psi task. Both questionnaires of 55 items (Groningen) and of 14 items (Amsterdam) can be found in the table in Appendix 1.

Analysis

To begin, and as a didactical element, the student experimenters had to analyze if, like in Bem's original experiment, anomalous interference of the random future with the present could be replicated. However, as mentioned before, our real goal of the study was to find questions predictive of performance in this retroactive priming task. In this analysis, for most of the questions a positive direction of the Spearman correlation with psi scores was expected. This expectation was based upon the previous findings in the research on individual differences in psi scoring.

TABLE 1
Mean Psi Score, Defined as Difference in Response Time in Milliseconds
between Incongruent and Congruent Condition,
Compared with the Expectation Value of 0

Study	Mean-Psi Score (millisecs)	t	df	*p*
Groningen	1.37	0.21	221	0.80
Amsterdam	-2.34	-0.153	60	0.88

Results are very similar with log-transformed data.

Secondly the supervisors realized that the data matrix with a column with random scores (under the null) and many columns with questionnaire scores was formally equivalent to data matrices obtained in other psi experiments that have been developed to check for 'excess correlation' (von Lucadou 2006). This extra (acausal) correlation is one of the predictions of GQT. The analysis of this prediction thus has theoretical impact and therefore it was decided post hoc to run this analysis. It requires the researcher to define a measure for this anomalous connectivity. Rather than using the subjective measure used by von Lucadou and others, we developed an objective measure.

Results

Overall Psi Scores

In spite of not being the focus of this experiment, the overall psi score, being the difference of the mean noncongruent and the mean congruent response times corrected for outliers as is standard in the package supplied by Bem, was compared with a chance expectation of 0 to assess an overall psi effect.

The effects claimed by Bem (2011) could not be replicated, as is clear from Table 1.

Correlations with Questionnaire Items

The exploration of questionnaire items that could potentially be used to select 'gifted' participants was the student's major goal of this study.

In the Groningen experiment, 4 questions out of 55 did correlate 'significantly' ($p \sim 0.05$; two-tailed) with the performance in the retro-priming task (see Table 2). Since these questions are dependent upon each

TABLE 2
Marginal and Significant Results of Correlations
between Psi Score and Response to Questions

Study	Number of significant relations	Questions	R	P (2–t)
Groningen	4 out of 55	G3. Sometimes I sense something that I am unable to explain with my normal perception (hearing, seeing, feeling, tasting, smelling).	−0.138	0.043
		G9. I believe in life after death.	0.139	0.041
		G16. I consider myself as being creative.	−0.185	0.006
		G23. I have had an experience in which I felt like I was outside my body.	−0.127	0.062
Amsterdam	4 out of 14	A1. Some humans are capable of lifting objects by means of mental power.	0.244	0.058
		A8. Some people have the unexplained faculty to predict the future.	0.290	0.024
		A13. People may have a visionary moment that can be used to predict the future.	0.280	0.020
		A14. The claim that people can predict the future is nonsense.	−0.241	0.064

other, there is no straightforward way to tell if this number of significant relations is significant in itself. This question is even more relevant with regard to the correlation matrix obtained in Amsterdam that also contained 4 significant correlations from the 14 correlations that were measured. We return to this issue in the discussion of the potential models for our results and the Connectivity analysis to check for these models.

Secondly these results look not very surprising because the questions do relate to aspects that have been found to be 'predictors' in previous research. For the Groningen data, this concerns most notably 'creativity'. This could give confidence, but such confidence is not warranted: The directions of the results of three of the four Groningen correlations are opposite to what one would expect on the basis of earlier research.
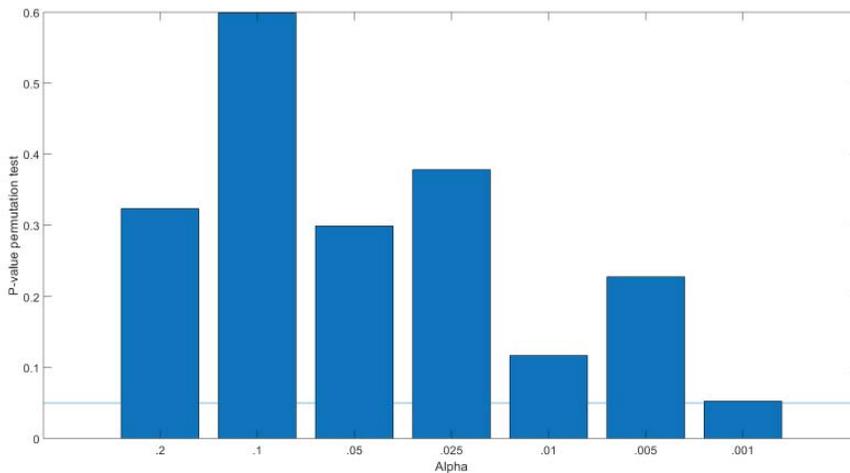
Figure 1. The *p*-value for the CMM (Correlation Matrix Method) analysis as a function of the criterion to count a correlation as significant.

Connectivity

In earlier research by von Lucadou (2006), an analytic approximation for the chance distribution of significant cells in correlation matrices was used. The approximation is required because of the dependencies among the responses to different questions. Grote (2017) did not use the analytic approximations but did *simulate* this distribution of the number of significant correlations in the correlation matrix, by multiple (10,000) random *permutations* of the actual psi scores of the different participants correlated to the question variables.

By comparing the obtained experimental chance distribution with the actual number of significant correlations obtained with the nonpermutated data, they assessed the probability of finding a correlation matrix that showed that amount of connectedness or more.

Several experiments were designed in the past to explicitly test the methods with the correlation matrix as the dependent variable (von Lucadou 2006), and more often than not such an effect was found. However, the statistical treatment has been criticized and data of the most recent experiments were re-analyzed using proper permutation methods (Grote 2017). This proper connectivity analysis does also support the psi hypothesis but the effect sizes are smaller than originally reported (a review article is in preparation).
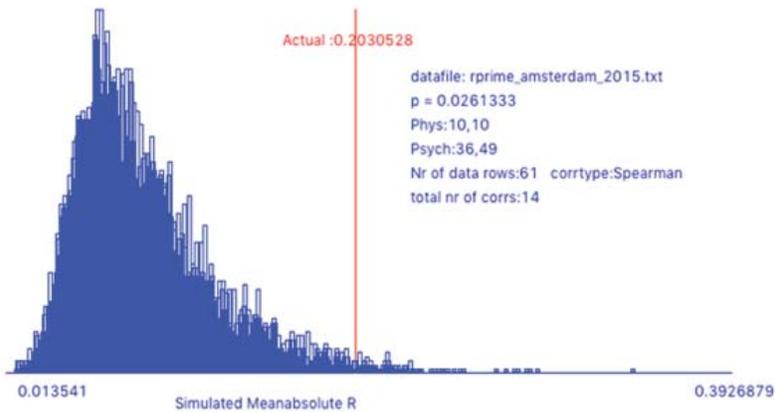
Figure 2. **Results of the 10,000 permutations analysis of the Amsterdam data. The red line represents the experimental data's real 'connectivity'. The histogram represents the control distribution of connectivity obtained by permutation techniques.**

### Removal of the Subjective Aspects of the Permutation-Based Connectivity Analyses

The choice of $p < 0.10$ as a criterion to call the correlation 'significant' is of course subjective. This can easily be seen if one uses different values for calling a correlation 'significant'. For instance in Figure 1 the results are dependent on this subjective choice, but for all values of $0.02 < p < 0.1$ the results of the analysis is significant.

Smaller values of p result in too few data points for a reasonable analysis.

Because of this subjective aspect, we developed a method that is not dependent on the arbitrary choice of a critical p-value. The key word here is *connectivity*. We explored two definitions for this concept. First we used the average of the absolute correlation values in the correlation matrix and secondly we looked at the average square of the correlation values in the correlation matrix. We used permutation techniques to construct the chance distributions for both of these dependent measures (Figure 2). (The software, 'connectivity-analysis', is available by downloading from the Heymans website or the Open Science Framework).

Permutation Analysis versus Monte Carlo Simulation

We also ran the analyses using random numbers rather than permutated psi scores. The results were virtually identical. We prefer the permutation
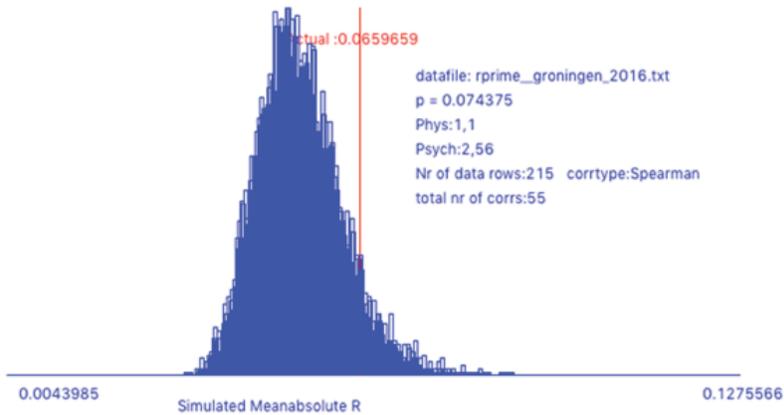
Figure 3. **Results of the 10,000 permutations analysis of the rPrime experiment run at the University of Groningen. The red line represents the 'actual connectivity' in the experiment. The histogram represents the control distribution of the explained variance.**

analysis because that analysis does keep certain aspects of the original distribution of the experimental psi scores.

Connectivity Analysis on an Identical Experiment in Groningen

According to the GQT, this anomalous surplus of connectedness in the correlation matrix should be replicable. Therefore, we also checked the dataset of the Groningen laboratory that participated in the global Bem experiment a year later. The correlation matrix analysis results were replicated in this dataset. Using the same permutation simulation to obtain the empirical chance distribution, it was found that the amount of connectedness in the correlation matrix of that experiment was marginally larger than can be expected by chance (see Figure 3, $p < 0.08$). These results can be seen as a *support* for the GQT. But of course for these conclusions to hold up, further replications are required.

Three New Measures for Connectivity

The requirement for a measure of connectivity is that it must be sensitive to correlations that are more extreme than we could expect by chance. We used the absolute value of the correlation, but one could also use the square of the correlation coefficient. The analyses using that definition of connectivity turned out to be less sensitive (see Table 3). Finally we used the standard

TABLE 3
p-**Values Obtained Using 3 New Different New Measures for Connectivity**

Measure	AMSTERDAM	GRONINGEN
<|R|>	0.018	0.074
<R*R>	0.029	0.175
sd(R)	0.025	0.235
Subjective #sig.R	0.013	0.599

Results for the old measure of number of significant correlation where significance is assumed if the correlation has a chance probability of 10% or less are shown for comparison.

deviation of the correlations. That measure gives results comparable to the measure based upon the mean square of the values of the correlations. For all the measures, the chance distribution was assessed by permutation techniques.

Discussion and Conclusion

The results of the evidential hypotheses did not replicate the original results from Bem (2011). It shows the reason why mainstream science is reluctant to accept the reality of these phenomena.

The research question from the students did not result in clear-cut answers. None of the reported correlation coefficients reached significance when corrected for multiple analysis. The most significant correlation was the one between 'creativity' and psi performance in the rPrime task. However, this correlation was negative, contrary to what is generally found and assumed in the field, namely that the more creative participants do better in psi experiments.

To see if this sign was correct, we did a number of post hoc checks on the integrity of the data. We checked the rPrime software, the entries by the students in the database of the responses to the questions, and the possible interference of the response times by the previous prime that appeared after the previous target.

After all these checks, we concluded that the significant correlations that were found had the proper sign. Therefore, these results may call into question earlier work on individual differences. Most notably regarding the factor 'creativity,' though, when corrected for multiple analyses the negative correlation is far from significant.

Procedural Differences

Confronted with this result, we identified some procedural aspects that might be different from other experiments that assessed individual differences. The major one is that the sessions were done in an experimental room with 2 or 3 half-open cubicles where the experiment was running. It has been suggested that this could have had an effect on some relations. For instance, the relationship with extraversion that has been reported (Bem & Honorton 1994) is thought to be strongly dependent on context. Extraverts would do better than introverts when tested in an open (lab) environment with more people. However, none of the questions dealt with extraversion.

Anomalous Connectivity in Regular Matrices

It should be remarked that in previous analyses of this kind the experiments were generally designed from the beginning to produce a data matrix that would be suitable for this approach. These PK experiments were generally rather complex to perform. The realization that simple data matrices from other than PK-experiments could also be used in this kind of connectivity analysis prompted us to use the data matrices obtained in the failed rPrime replications. The only requirement for the matrix is that at least one column contains random data under the null. This is generally the column with psi performance data. In fact in each psi experiment where personality variables are measured, such an analysis is possible. The software for the Connectivity analysis is able to handle most matrices.

Physical Theories

There are two theories rooted in physics that could be argued to account for elusive effects.

    CIRTS. In the first, consciousness-induced restoration of time symmetry (CIRTS) (Bierman 2010), all psi is modeled by retrocausation (or, more accurately, time symmetry). However, this retrocausation cannot be used to create a so-called *closed time loop* paradox (like the grandfather paradox in time travel). This paradox occurs when 'future' information is used in the past to change the course of events that resulted in that future information. This limiting principle in the theory fits with ideas by Novikov (Novikov 1992) and Echeverria (Echeverria, Klinkhammer, & Thorne 1991). Time travel to the past is possible but not to everywhere in space–time. It will be impossible to change the course of events in such a way that the future changes. The result of this limiting principle may look like the classic Trickster chaos. But there should be some internal structure; these 'Trickster' moments should be more probable in contexts where closed time

loops that may result in paradoxes that could be produced. In principle this is a testable hypothesis.

Generalized Quantum Theory. The second theory is called the Generalized Quantum Theory (GQT). Here the limiting principle is that the anomalous correlation may never be used as a *signal* (Atmanspacher, Römer, & Walach 2002, von Lucadou & Römer 2007). That is, it will be impossible to manipulate at will on one side of the 'communication channel' thereby forcing the other side to be encoded at will. This is an analogy to the quantum nonlocal correlations that disappear if one sets up the experiment in such a way that the correlation may be used as a signal (for instance to earn money). But rather than being totally capricious, the built-in 'Trickster' in this theoretical approach is assumed to become active only when the nonlocal character of the correlation is threatened.

A well-known 'practical' prediction of the GQT is that although upon replication, according to the theory, the anomalous correlation will disappear in the (correlation matrix) cell that one intended to use as a signal cell, extra correlations will pop up in other (unexpected) cells. The hypothesis is then that the correlation matrix that we found contains more extreme correlation values (independent of sign) than can be expected by chance. As argued in the Results section, there is no straightforward analytic statistical technique to test this hypothesis because the cells are dependent and therefore the null distribution of the connectivity in the correlation matrix is unknown. We solved this problem by using permutation techniques to construct the chance distribution of two variables that could be used as operationalization of 'Connectivity'. The use of the average *absolute* correlation coefficient turns out to be more sensitive than the use of the average of the square of the correlation matrix. This could imply that cells with weaker correlations are more apt in assimilating the supposed anomalous nonlocal correlation than cells that have more extreme values, and this finding if replicated may give further theoretical suggestions.

To summarize: The experiment was set up for the student experimenters as an exploration into possible individual differences in performance in a psi task. Overall performance was at chance level. Some correlations between the performance and several predictors were larger than can be expected by chance, but the direction of the correlation for some questions went against the expectation. However, when corrected for multiple analyses all these correlations become nonsignificant.

The major finding from these experiments comes from a post-hoc analysis suggested by the apparent decline from Bem's results and the Generalized Quantum Theory that 'explains' this decline and predicts that in case of decline the anomalous correlations should 'go' to other cells

in the correlation matrix. Although the whole study was explorative and hadn't been preregistered, we still believe that this finding may be different from the chance findings that always pop up in explorative research. Most important is the fact that we initially assessed this effect marginally in the Groningen data, then requested the totally independent Amsterdam data and could replicate this result. We cautiously conclude that this finding can be seen as a support of the Generalized Quantum Theory.

Future Developments

The use of the term *connectivity* may turn out to be misleading. At least in the Amsterdam data, we found a comparable effect using the standard deviation of the values of the correlation coefficients. This could be interpreted as an increase of variance rather than connectivity. Currently we are running simulations with several models for how the data may be affected in such a way that we get the same pattern of results as we found using the three different new measures. These simulations may therefore give the theoretical details of how the concept of nonlocal correlation as used in the GQT may interfere with normal causal correlations.

Acknowledgments

References Cited

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics*, *32*(3):379–406.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3):407–425.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*(1):4–18.

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research*, *4:*1118.

Bierman, D. J. (2010). Consciousness induced restoration of time symmetry (CIRTS): A psychophysical theoretical perspective. *The Journal of Parapsychology*, *74*(2):273–299.

Bierman, D. J., Spottiswoode, J. P., & Bijl, A. (2016). Testing for questionable research practices in a meta-analysis: An example from experimental parapsychology. *PLOS One*, *11*(5): e0153049. DOI: 10.1371/journal.pone.0153049

Boer, R. de, & Bierman, D. J. (2006). The roots of paranormal belief: Divergent associations or real paranormal experiences? *Proceedings of the 49th Annual Convention of the*

Parapsychological Association, edited by Christine Simmonds-Moore, Stockholm, 4–6 August, 2006, pp. 283–298.

Echeverria, F., Klinkhammer, G., & Thorne, K. S. (1991). Billiard balls in wormhole spacetimes with closed timelike curves: Classical theory. *Physical Review D, 44*(4):1077–1099.

Grote, H. (2017). Multiple-analysis correlation study between human psychological variables and binary random events. *Journal of Scientific Exploration*, *31*(2):231–254.

Honorton, C. (1997). The ganzfeld Novice: Four predictors of initial ESP performance. *The Journal of Parapsychology*, *61*(2):156–158.

Jolij, J., & Bierman, D. J. (2017). Testing the potential paradoxes in "retrocausal" phenomena. *AIP Conference Proceedings, Quantum Retrocausation III*, *1841*(1):030003.
DOI: 10.1063/1.4982774

Lucadou, W. von (2006). Self-organization of temporal structures—A possible solution for the intervention problem. In *Frontiers of Time. Retrocausation—Experiment and Theory* edited by Daniel P. Sheehan, *AIP Conference Proceedings 863*, Melville, NY: American Institute of Physics, pp. 293–315. DOI: 10.1063/1.2388760

Lucadou, W. von, & Romer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *Journal of Consciousness Studies*, *14*(4):50–74.

Novikov, I. D. (1992). Time machine and self-consistent evolution in problems with self-interaction. *Physical Review D, 45*(6):1989–1994.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6):528–530. DOI: 10.1177/1745691612465253

Schmeidler, G. R., & McConnell, R. A. (1958). *ESP and Personality Patterns* (New Haven, CT: Yale University Press).

Traxler, M. J., Foss, D. J., Podali, R., & Zirnstein, M. (2012). Feeling the past: The absence of experimental evidence for anomalous retroactive influences on text processing. *Memory & Cognition*, *40*(8):1366–1372.

APPENDIX 1
Full List of Questions and Correlations with Psi Score

N in the Groningen data ranges from 213 to 215. N in the Amsterdam data is 61.

nr	Question	R	p
G1	I see beauty in things where others might not.	−.058	.398
G2	When getting angry, I direct my aggression toward others instead of keeping it to myself.	.030	.666
G3	**Sometimes I sense something that I am unable to explain with my normal perception (hearing, seeing, feeling, tasting, smelling).**	**−.138[*]**	**.043**
G4	At least once I have found myself in a state of altered consciousness (which is a temporary change in one's normal mental state without being considered unconscious, so not while dreaming, daydreaming, meditating, or being hypnotized).	−.030	.664
G5	I believe I am able to see things that are in the future or happening somewhere else.	−.100	.144
G6	I often have visible signs of nervousness such as sweaty palms or shaky hands.	−.029	.673
G7	When making decisions, I often rely on my gut feeling instead of logical thinking.	.075	.274
G8	I believe that it is possible to sense things not with your physical senses (feeling, hearing, seeing, etc.) but with your mind, as if it is "the sixth sense".	−.074	.282
G9	**I believe in life after death.**	**.139[*]**	**.041**

nr	Question	R	p
G10	I am confident in succeeding in all parts of this experiment.	.037	.586
G11	I'm usually aware of internal processes like thoughts, feelings, and bodily sensations.	−.017	.805
G12	I believe in paranormal phenomena (e.g., witchcraft, telepathy, existence of ghosts).	−.044	.519
G13	I usually feel comfortable in testing situations.	−.055	.424
G14	I frequently have self-critical thoughts.	−.079	.250
G15	When in trouble, I actively try to improve the situation.	−.038	.583
G16	**I consider myself as being creative.**	**−.185****	**.006**
G17	I have had an out-of-body experience.	−.069	.314
G18	I believe it is better to think things out than to get angry.	−.104	.131
G19	I have been successfully hypnotized on a previous occasion.	−.075	.275
G20	I consider myself as an outgoing person.	−.088	.202
G21	I am never bored because I start fantasizing when things get boring.	−.077	.261
G22	When reflecting on my personality, I feel that I have many qualities which are traditionally regarded as feminine.	−.047	.492
G23	I've had an experience in which it felt like I was outside my body	−.127	.062
G24	I often do things without thinking through the possible outcomes.	.022	.753
G25	I feel it is important for me to act as an independent person.	−.029	.674
G26	I practice ways of mental control, such as meditation, self-hypnosis, autogenic training, etc.	−.110	.110
G27	I often feel that I have little influence over the things that happen to me.	.081	.236
G28	Sometimes I am aware of the fact that I am dreaming.	−.001	.984
G29	I have experienced or am experiencing mental health conditions.	.095	.164
G30	I get stressed out easily.	.034	.618
G31	I am open to new experiences.	−.046	.503
G32	I have participated in research about whether I can anticipate or sense the future.	−.039	.568
G33	I make rational decisions.	.029	.677
G34	At least once I have experienced telepathy (which is transmitting information to someone without any form of interaction).	−.086	.208
G35	I feel relaxed in new situations.	−.012	.863
G36	Religion is an important part of my life.	.108	.114
G37	I am comfortable taking risks even though I might be wrong.	−.055	.422
G38	I have a lot of sex appeal.	−.041	.555
G39	I have an inner aspiration to go beyond previous limits in different fields (physical, mental, spiritual).	−.087	.204

nr	Question	R	p
G40	I like to find myself in situations which make my heart beat faster.	.009	.892
G41	I smile at strangers whenever it is appropriate.	−.081	.238
G42	I am easily bothered by stimuli in my environment, e.g., noises or chaotic scenes.	.110	.108
G43	The norms and values that influence my goals, plans, actions, thoughts, and feelings, are to a great extent based on my spiritual beliefs.	−.057	.404
G44	In familiar situations, I enjoy the unexpected.	−.023	.743
G45	At least once I have seen something that was happening in the future or somewhere else.	−.090	.187
G46	I feel I have achieved something in my life.	.081	.235
G47	When asked to choose a number, I tend to choose my lucky number.	−.056	.417
G48	When I am in pain, I feel I am being slower.	.041	.548
G49	I believe that when the person I love the most got seriously ill or died, I would feel or know it somehow.	.077	.262
G50	Someone has tried to take something directly from me by using violence or threat of violence.	−.022	.749
G51	I often deceive people.	.054	.434
G52	I am interested in reading books and articles on psychic phenomena.	−.111	.103
G53	I use meditation techniques.	−.075	.274
G54	I consider myself an artistic person.	−.074	.281
G55	I believe I have telepathic abilities (which is the ability to transmit information to someone without any form of interaction).	−.085	.215
A1	Some humans are capable of lifting objects by means of mental power.	.244	.058
A2	PK, the movement of objects by means of psychic influence, does really exist.	.099	.448
A3	Thoughts can move objects.	.133	.305
A4	It is impossible to read someone's mind.	.160	.218
A5	Astrology can predict the future.	.232	.072
A6	A horoscope may foretell the future.	.218	.092
A7	Some psychics are able to accurately predict the future.	.224	.083
A8	**Some people have the unexplained faculty to predict the future**.	**.290**[*]	**.024**
A9	Events in the future may influence my behaviour and decisions at this moment.	.228	.077
A10	My intuition enables me to feel the future.	.201	.121
A11	It is possible that by unknown means a mental picture can be transferred.	.196	.131
A12	It is possible to mentally influence some other person by thoughts alone.	.093	.476
A13	**People may have a visionary moment that can be used to predict the future.**	**.284**[*]	**.026**
A14	The claim that people can predict the future is nonsense.	−.241	.061

APPENDIX 2
Full Text Used for Selection of Student Experimenters

Can future events have an effect in the present? According to a controversial paper by Bem (2011), it can. With this thesis, you can replicate Bem's famous experiment. This project is part of a larger international replication effort. Each student in this group will run his/her own experiment. You will learn through doing the following:

1. Adjust a foreign experiment to local conditions and replicate a so-called 'retroactive priming' experiment with 20 subjects. The experiment consists of a computerized priming task with English words (that have to be adapted for Dutch/German-speaking subjects) and a questionnaire that also has to be translated from English. This work will be split up among the 5 students in the group (group effort).
2. Formulate a personal hypothesis and pre-register this (personal effort).
3. Expand the questionnaire with new items that are supposed to be predictors of the subject's performance in the priming task (group effort).
4. Introduce each of the 20 subjects to the experiment, keep a logbook, and debrief the subject at the end of the experiment.
5. Do a simple differential test comparing the priming response times in two conditions (congruent versus noncongruent).
6. Analyze item scores of the questionnaire with the priming performance scores.
7. Write a so-called extended abstract about your own experiment.