Open Access

Do You Know Who is Calling? Experiments on Anomalous Cognition in Phone Call Receivers

Stefan Schmidt[1,*], Devi Erath[1], Viliana Ivanova[1] and Harald Walach[2]

[1]*Department of Environmental Health Sciences, University Medical Center Freiburg, Germany and* [2]*School of Social Sciences, University of Northampton, UK*

Abstract: Many people report that they know in advance who is on the phone when the telephone is ringing. Sheldrake and Smart [1, 2] conducted experiments where participants had to determine which one of four possible callers is on the phone while the telephone was still ringing. They report highly significant hit rates that cannot be explained by conventional theories.

We attempted to replicate these findings in a series of three experiments. In study one, 21 participants were asked to identify the callers of 20 phone calls each. Overall 26.7 % were identified correctly (mean chance expectation 25%, *ns*). In a second study a pre-selection test was introduced in a different experimental setting. Eight participants identified 30% of the calls correctly (p = .15). However one of the participants recognized 10 out of 20 calls correctly (p = .014). We conducted a third study with only this participant. In an additional 60 trials she could identify 24 callers correctly (p = .007). We conclude that we could not find any anomalous cognition effect in self-selected samples. But our data also strongly suggest that there are a few participants who are able to score reliably and repeatedly above chance.

Key Words: Anomalous cognition, telephone experiments, ecological validity, telepathy.

INTRODUCTION

Many people report strange experiences when called by or calling others on the telephone. There are reports of situations where a person intends to call another person and in the very moment when s/he wants to lift the receiver this other person calls in. Or the telephone rings and the called persons have an intuitive knowledge of who is calling when they answer the phone. Some of these and other similar experiences seem to be beyond any reasonable explanations, such as expected phone calls or regular calling schedules, and leave the affected persons stunned.

Rupert Sheldrake and colleagues conducted two surveys on this phenomenon in California and the UK [3, 4]. Amongst other questions, they asked both populations: "Have you ever heard the telephone ring or picked up the telephone and known who was on the other end without possible cue, before they have spoken?" In Lancashire, 49% of 200 persons asked answered "yes" while in California 47% answered "yes". While some of these experiences might be explained by coincidences, many people tend to believe in unconventional explanations because their experiences were either too extraordinary to attribute it to a coincidence or just made sense in a perfect way. Thus Sheldrake calls these phenomena "… the commonest kind of apparent telepathy in the modern world." [1]. They next set out to back up this claim with empirical proof by setting up a simple experiment.

In this experiment [1] a participant is called by one of four pre-selected callers. The callers are selected by a random process and there is no conventional possibility of knowing in advance who the caller will be. The participant guesses the name of the caller before answering the phone. There is a mean chance expectation (MCE) of p = .25 of making a correct guess. If the above reported phenomena are just due to chance, unconscious expectancies, selective memory of these events or other conventional explanations, then this mean chance expectation should be found if the experiment is conducted thoroughly and there is no sensory leakage of the relevant information. But if on the other hand the intuitive knowledge on the side of the callee is due to some anomalous cognition (whatever this might be), then the MCE should be exceeded in such an experiment.

In their first publication, Sheldrake and Smart [1] report on several series of 571 trials in total, with 63 different participants. Four different experimental protocols were applied in these series, but the basic task to identify one of four potential callers was maintained. Overall there were 231 correct guesses (hit rate 40 %, $p = 4 \times 10^{-16}$)[1]. Thus the anomalous cognition hypothesis was apparently confirmed by an extraordinarily strong effect. However, on close examination this study showed several limitations. One weakness was that the experiment was rather uncontrolled. Participants waited at home for about an hour for the target person to call. After the call the experimenter phoned the participants and asked what they had guessed. In some trials this was

*Address correspondence to this author at the Department of Environmental Health Sciences, University Medical Center Freiburg, Breisacherstr. 115b, D-79106 Freiburg i. Br., Germany;
E-mail: stefan.schmidt@uniklinik-freiburg.de

[1] A recalculation by us yielded a slightly smaller but nevertheless highly significant p-value of 1.7×10^{-14}.

1874-3501/09 2009 Bentham Open

counterchecked with the respective caller but, nevertheless, this procedure allows for a whole range of manipulations on the side of the participant. A second limitation is even more serious. When the telephone rang the participants first lifted the receiver and then spoke the name of the guessed person aloud into the receiver. This procedure of defining the guess after lifting the receiver allows for sensory leakage. A typical clicking, a humming sound, or some background noise from the caller might give a relevant clue.

They conducted a second study with much tighter controls [2]. In this experiment only selected participants were employed. Each participant had to reach a minimum success rate in a pretest with ten trials. Four participants remained for a total of 271 trials (183 trials by only one participant). In these trials the subjects were videotaped while awaiting the call. They spoke their guess into the video camera before lifting the receiver and thus the data of this study can be considered the result of a valid and well controlled examination of the hypothesis under examination. Overall there were 122 correct guesses (hit rate 45%, $p = 10^{-12}$, Effect size h=0.42), and the authors conclude this to be proof of an unconventional information transfer. They also tested a second hypothesis on the difference between callers who are known to the participants and those who are not. Therefore, in 175 trials only two of the four callers were familiar to the participants while the other two were recruited by the experimenter. Sheldrake and Smart hypothesize that only calls by familiar callers will result in an increased hit rate, while the unfamiliar callers will be identified no better than chance. They report hit rates of 61% for familiar callers and 20% for unfamiliar callers with an apparently significant difference ($p = 8 \times 10^{-8}$). However the statistical procedure they applied was unsatisfactory. The data were not corrected for a strong response bias towards guessing more familiar callers. When corrected for this bias, the hit rates reduce to 46% for familiar callers and 37% for unfamiliar callers, respectively [5]. The difference between these hit rates was no longer significant ($p = .32$). However, in a reply to this critique Rupert Sheldrake [6], showed that the data reach significance again if a different statistical analysis is applied (randomization test, $p = .0002$).

Lobach & Bierman [7] replicated the experiment with an additional independent variable. Local sidereal time (LST), an indicator of the earth's orientation with respect to the stellar background, has been shown to be related to effect sizes in experiments on anomalous cognition [8]. A reanalysis of Sheldrake's findings revealed that most of his trials were conducted at a time window known as LST peak time (i.e. 13.30 h LST). Thus Lobach & Bierman compared 107 trials scheduled at peak time with 107 trials scheduled at non peak times in their telephone experiments. They obtained a hit rate of 34.6% at peak time and 25.2 % in the non-peak condition. The overall hit rate of 29.4% just reached significance ($p = .05$).

Whereas the Lobach & Bierman study appears to be a straightforward test of the relevant hypotheses, the Sheldrake studies have a rather exploratory nature, with frequent changes in the design and only partially prespecified numbers of trials. We thus set out to conduct a set of more rigorous conceptual replications of this experiment. Three studies were conducted. In study one participants were invited to an office for two consecutive sessions of 10 calls. In study two and three we visited the participants at home for several sessions of 5 calls. Overall we conducted 557 trials with 29 different participants. All trials were videotaped. All studies tested two hypotheses: (1) Callees will be able to identify the incoming calls significantly better than chance (2). Callers familiar to the callee will be identified significantly better than unfamiliar callers.

STUDY ONE

Methods

We planned to conduct a study with a preplanned number of 400 trials and 20 participants on the basis of a power analysis of the Sheldrake studies. Participants were invited for two sessions of 10 trials each. Twenty-one participants were recruited through newspaper advertisement, press reports and leaflets. Nineteen participants contributed 20 trials in two sessions of ten. Two participants were not able to participate in a second session and contributed ten trials only. Seventeen participants (81%) were female, and the mean age of all participants was 39.8 years (SD = 11.3, range 22-62). Participants were paid 20 Euros each for their participation. The study was organized, conducted, and analyzed by Susanne Müller [9]. The experiment took place in the office space of a youth counseling organization in Freiburg, Germany in 2003. All sessions were videotaped. The telephone for the incoming calls had no display for caller identification. Randomization and scheduling of the callers took place in a second office located in a distance of approximately 3km.

Each participant provided the names of two familiar callers willing to participate in this study. These familiar callers had to be available by telephone during the experiment, either in their homes, at work, or elsewhere if they had a mobile phone. Two unfamiliar callers were provided by a second experimenter in a distant office. For the randomization of the callers the Marsaglia-Zaman random number generator was employed [10]. As soon as the participant arrived in the office the second experimenter scheduled ten incoming calls for this session. The callers were then called by the experimenter by phone and informed of the telephone number to call and the calling time. During the next 90 minutes the participant received ten calls, approximately one every ten minutes. When the telephone rang the participant had to guess which one of the four possible callers was on the phone. They spoke their guess into the video camera, and the experimenter recorded the guess on a session sheet. Then the experimenter lifted the receiver and checked the caller. This information was fed back to the participant. Between calls the participants were free either to read some of the magazines provided containing content unrelated to the experiment or to chat with the experimenter. All procedures, including the analysis strategies, were preestablished in a research protocol that was deposited with a neutral person before the start of the data collection.

Statistics

The mean chance expectation for each single trial is $p = .25$ and all four targets within each trial have the same probability independently of the prior trials. For Hypothesis 1, the likelihood of X hits within N trials can be calculated by the formula for the critical ratio:

(*CR*): $CR = (X - Np)/sqrt(Npq)$ (with $q=1-p$) [11].

CR is equivalent to a z-score and normally distributed. The *CR* can be calculated for any given number of trials independently whether they are from one or several participants. This analysis was prespecified for the evaluation of hypothesis 1. The test of significance was one-tailed. Another way of assessing the results is a binominal test assessing the likelihood of obtaining *X* or more hits out of *N* trials (one-tailed), when the probability of obtaining a hit is $p = .25$. We calculated the effect size Cohen's *h* for the comparison between the empirical result and an expected distribution using the formula:

$$h = (2 \arcsin \sqrt{p_{exp}}) - (2 \arcsin \sqrt{p_{mce}})$$

where p_{exp} stands for the hit rate and p_{mce} for the MCE [12, p. 181]. For Hypothesis 2 we calculated a χ^2-test for the distribution of the data in 2 x 2 table with the rows hits (yes – no) and the columns calls (familiar – unfamiliar). This test returns a *p*-value for the likelihood that the distribution of the hits is independent of expectancy for a familiar or unfamiliar caller.

Results

We obtained 397 valid trials. Data from three trials were excluded from the analysis because the call arrived too late. Two-hundred-and-one calls came from familiar callers and 196 from unfamiliar callers. As expected, this ratio shows that the randomization of callers (familiar/unfamiliar) was unbiased ($p = .84$, binominal test, two-tailed).

On the other hand, the distribution of the guesses for these four callers by the participants was not random. In most cases participants guessed the caller to be a familiar one, with an additional preference for the first familiar caller they had named. The numbers were: guesses for familiar caller 1: 131; guesses for familiar caller 2: 111; guesses for unfamiliar caller 1: 96; guesses for unfamiliar caller 2: 59. Overall these are 242 guesses for familiar callers compared to 155 for unfamiliar ones. This distribution is significantly different from chance expectation ($p = .00001$, binominal test, two-tailed).

Hypothesis 1: The participants identified 106 of the 397 calls correctly. This results in a hit rate of 26.7 % (MCE = 25%) and a critical ratio of z = .78 (*ns*), effect size $h=0.04$. The binominal test yields a p-value of .23. Thus, hypothesis 1 was not confirmed. Hypothesis 2: From the 106 correct guesses 67 (63.2%) were with familiar callers and 39 (36.8%) with unfamiliar callers. Based on the participants' guesses for familiar or unfamiliar callers the hit rates for familiar callers is 27.7 % (*ns*) and for unfamiliar callers 25.2 % (*ns*) respectively. The χ^2-test on the difference results in a non-significant *p*-value of .58 ($\chi^2 = 0.31$, $df=1$) and thus hypothesis two was not confirmed.

Discussion

In this more tightly controlled study, we failed to replicate the positive findings reported by Rupert Sheldrake. There was neither a hit rate larger than expected by chance, nor did we find any significant difference between the hit rates for familiar and unfamiliar callers. In short, we find no evidence for anomalous cognition in the ability to identify telephone callers.

However our study was different from the original studies in several respects: (i) The experiment took place in a room provided by us rather than in the participants' homes; (ii) The experimenter was present during the session and was the one to lift the telephone receiver; and (iii) we used a self-selected sample rather than restricting it to high scorers on a pretest.

To determine whether the more private home setting is needed or whether perhaps only selected individuals are able to identify the callers better than chance, we conducted a second study with a preselected sample and higher ecological validity.

STUDY TWO

Methods

We planned to conduct a study with 100 trials in an adaptive design. Each person interested in participating in the experiment was invited to conduct a pre-test. When they passed the pre-test successfully, they were invited to participate in two sessions with five calls each. Participants who reached four or more hits ($p = .22$) in these initial ten sessions were then invited to participate in another ten sessions and so on. Using this strategy we recruited participants until the pre-planned 100 trials were completed. With this adaptive design we could continue to work with individuals with high hit rates without running into the problem of optional stopping [13]. Optional stopping is a procedure capable of creating artifacts in guessing experiments by stopping data collection once the average hit rate is just above chance. In order to avoid this, the number of trials and recruitment strategies must be determined before the start of the data collection, as was done here.

Participants were recruited by advertisements, through the internet, and among friends and relatives of the experimenter. The pre-test consisted of an email experiment similar in design to the telephone experiment provided by Rupert Sheldrake on his homepage. At a pre-planned time a participant and 2-4 of their friends had to log on to this webpage from computers at different locations. The system then randomly selected one of the friends as a target person and asked them to send an email to the participant via the remote system. The participants had to guess the sender before they received the mail but after the mail was sent. This procedure was repeated for several trials. The participants were asked to conduct this pretest by themselves and to provide the print out of the overall results as a proof. For the first part of this study the criterion was set to 4 or more hits out of 10 trials. There is a probability of 22% of passing this test by chance alone. Six out of 25 (24%) participants reached either 4 hits (5 persons) or 5 hits (1 person) and were invited to take part in the controlled telephone experiment. In this second stage each of them participated in 10 trials in which they tried to identify a telephone caller. Only one of them reached the criterion of 4 correctly identified calls, but this participant was not interested in continuing with the experiment. Due to external constraints (recruitment took longer than expected, deadline for submission of thesis was reached) the experiment had to be stopped after these initial 60 trials. The experiment was organized and conducted in 2005/06 by the second author, DE [14] in Freiburg, Germany. After a break

of 12 months the third author, VI, continued the study in Freiburg in the Spring of 2007, providing the missing 40 trials. At that time the internet based email experiment was no longer available in its previous form. The new version required 3 friends and had only 9 trials. We set the criterion to 5 correctly identified emails, thereby setting the criterion (14%) as close to the original as possible. During the recruitment of new participants, this test was also replaced by a still newer version on the Sheldrake webpage. This changed version required three friends with only six trials. Here we set the criterion to 4 or more hits out of 6 with a probability of 10%. Two more participants were recruited for the remaining 40 trials. The first one achieved 7 correct guesses within 9 trials of the email test. In the telephone experiment he had 5 hits out of 10 trials and was invited to continue with a second set of ten trials. Here he scored 2 hits out of ten and failed to qualify for another set. The last participant had 5 correct guesses out of 6 in the email test. Within the first 10 trials of the telephone experiment she scored 5 hits and was invited to continue. She again scored 5 hits out of 10 but at that point the pre-specified 100 trials were completed. We decided to continue with this participant in a separate study (reported below). Overall 8 participants completed 100 trials in study two. The average age was 32 years, 6 were female.

This telephone experiment took place at the participants´ homes. Participants were asked to place themselves in a comfortable position in front of their telephone and to remain there during the trial. A video camera was set up and both the participant and the telephone were continuously recorded. Videotapes were inspected later on for the occurrence of any event which could invalidate the data. If participants had a telephone with a display showing incoming callers the experimenter covered this display with opaque tape. Mobile phones had to be switched off and placed on a table or chair within the range of the video camera. Once the participant received the instructions and the video camera was turned on, the experimenter left the participant's home. The experimenter selected the caller for the next trial by throwing a 6-sided die. The numbers from 1 to 4 were assigned to the four callers, and in cases of throwing a five or six the procedure was repeated. We used only high quality dice manufactured for use in casinos. Two of the potential callers were provided by the participant (familiar callers); the other two were provided by the experimenter (unfamiliar callers). Next the exact calling time within a range of 10 minutes was determined by throwing the die again. The experimenter then called the selected caller and told him or her which number to call at which time. At the designated time the selected caller called the participant. Once the participants' phone rang they spoke their guess of which person might be calling into the video camera, and also noted it on a protocol sheet. They also rated their confidence in their answer ('very sure', 'not so sure' or 'just guessed'). Next they picked up the phone and spoke shortly with the caller. The experimenter then returned to the participant's home, took note of the results and switched off the camera. After a short break the same procedures were repeated for the next trial. Overall a session consisted of five trials. A second session with another five trials was conducted on a different day. When all 10 trials were completed, participants were paid 30 Euros. Again all procedures including the analytical strategies were preestablished in a research protocol that was deposited with a neutral person before the start of the data collection.

Results

We used the same analytical strategies and hypotheses as in study one.

Hypothesis 1: The participants identified 30 of the 100 calls correctly. This results in a hit rate of 30 % (*MCE*=25%) and a critical ratio of $z = 1.16$ ($p = .12$, one tailed). The binominal test yields a p-value of .15, effect size $h = 0.11$. Thus, hypothesis 1 was not confirmed.

Hypothesis 2: From the 30 correct guesses 20 were with familiar callers and 10 with unfamiliar ones. Familiar callers called in 59 trials, unfamiliar in 41 cases ($p = .09$, binominal, two-tailed). In 59 cases, the participants guessed the caller to be familiar; in 41 cases they guessed the caller to be unfamiliar. The hit rates for familiar callers was 33.9% ($z = 1.58$, $p = .057$, binominal test $p = .08$); for unfamiliar callers, it was 24.4 % (*ns*), respectively. The χ^2-test on the difference results in a non-significant $p = .31$ ($\chi^2 = 1.04$, $df = 1$), and thus hypothesis two was not confirmed. Participants rated the confidence of their guesses on a scale with three possibilities. The respective hit rates were 42% (10 hits out of 24) for 'very sure', 28% (16 hits out of 42) for 'not so sure' and 22% (4 hits of 18) for 'just guessed'. The difference in the respective z-scores according to the formula $z_{diff} = (z_1 - z_2)/\sqrt{2}$ for the largest difference between 'very sure' and 'just guessed' resulted in $z = 1.53$ ($p = .06$, one-tailed). The last of the eight participants had an individual hit rate of 50% in 20 trials ($z = 2.58$, $p < .01$, one-tailed).

Discussion

Study 2 differed from study 1 in two ways. First, the experimental situation was designed to be similar to a real life situation, with the participants being in their own homes with their own telephone and without anybody else present. While more time-consuming, this had the advantage of higher ecological validity and of being a direct replication of the original studies by Sheldrake and Smart (2003a, 2003b). However, these changes still did not yield significant results. There was a slight excess of hits above the MCE but this remained within chance expectations. The other change was the introduction of a screening test to identify potential high-scoring participants. Owing to changes in the web-based test, we had to use different thresholds for including participants in the main study. These ranged from $p = .10$ to $p = .22$. Out of 27 persons participating in this screening test, 8 (29.6 %) reached the required threshold. Thus it is likely that most of the participants selected by this procedure were selected due to a type 2 error and not by an ability to detect the incoming emails correctly through anomalous cognition. The non-significant findings in the main telephone experiment support this view. Thus if it is really true that there is a small percentage of the population able to score repeatedly better than chance in these trials, then it may be that the criterion for the screening test was insufficiently strict for selecting participants. However, one of our participants scored 5 hits out of a possible 6 (binominal $p = .02$) in the screening test, followed by five hits in ten telephone trials (binomial $p = .08$), followed by another five hits out of ten in the second series. In order to find out whether these results were

Table 1. Raw Data of the 60 Trials of Study 3, Correct Guesses are Printed in Bold

		Caller				Total	% Right
		Familiar 1	**Familiar 2**	**Unfamiliar 1**	**Unfamiliar 2**		
Guess	Familiar 1	**12**	6	2	3	23	52
	Familiar 2	4	**6**	4	6	20	30
	Unfamiliar 1	3	4	**4**	0	11	36
	Unfamiliar 2	0	2	2	**2**	6	33
	Total	19	18	12	11	60	40

due to chance or could be replicated we conducted a third study with only this participant.

STUDY THREE

Method

As noted above, study three is an exploratory study using the most successful participant from study two[2]. The design was exactly the same as in study two. A series of 10 calls consisted of two sessions with five calls each. We did not preplan a fixed number of trials but asked the participant at the end of each series of ten calls whether she would like to continue with another series of ten calls. In all, we conducted 6 series with 60 calls within 5 weeks. The experimenter was the same experimenter as in the second part of study two. All sessions were videotaped. During the last two trials of the first part of series 3 the video-camera broke down. Thus two trials were not recorded. One of them was a hit.

Results

Twenty-four of the 60 incoming calls could be identified correctly. This is a hit rate of 40 % (MCE 25%) and yields a critical ratio of $z = 2.68$ ($p<.01$, binominal $p<.01$), effect size $h = 0.32$. The raw data are displayed in Table **1**. The 60 trials were completed within 6 series of 10 calls. The participant scored within the six series of 10 calls 5, 3, 4, 4, 4, and 4 hits. Out of 60 calls, 37 came from familiar callers ($p = .09$, binominal, two-tailed). The participant guessed 43 of the 60 calls to come from familiar callers. This is a significant response bias ($p = .001$, binominal, two-tailed). If the first 20 trials from study two are added to this data, one can see this participant's overall performance. The overall hit rate results in 42.5 % (34 hits in 80 trials, $z = 3.6$, $p = .00015$, $h = 0.37$). Familiar callers were identified with a hit rate of 42% ($z = 2.55$, $p = .005$, binominal test $p = .01$) and unfamiliar callers with a hit rate of 35% respectively ($z = 0.98$, *ns*). The χ^2-test on the difference is non-significant ($\chi^2 = 0.22$, $df = 1$). If the first 20 trials are added, then familiar callers were identified with a hit rate of 46% ($z = 3.6$. $p = .0002$) compared to 35% ($z = 1.1$, *ns*.) for unfamiliar ones. The χ^2-test shows no significant difference but the difference of the two respective z-scores (see formula above) just reaches significance ($z_{dift} = 1.75$, $p = .04$, one tailed). The confidence rating did not show any correlation with the experimental results. The participant rated her confidence in 20 trials as 'very sure' and in 40 as 'not so sure', the respective hit rates were 45% and 37.5%. She never checked the 'just guessed' alternative.

Discussion

The most successful participant from study two continued to score slightly above chance in another 6 series of telephone experiments. Overall she reached a significant hit rate with a medium effect size and the results from study three replicate the results for this particular participant from study two. Could a conventional, non-anomalous explanation account for these results?

There are two alternative explanations we can think of: sensory cueing and fraud. Sensory cueing is quite unlikely because the design of the experiment strictly controlled for this possibility. The caller was only determined after the participant was already monitored by the video camera and thus every sensory cue should be detected by an inspection of the video tapes. We have done so but could not find any hint of information leakage. Fraud is much more difficult to rule out, but because the video camera monitors the participant continuously, fraud could only take place if the experimenter and the participant collude to do so. We see this as a very unlikely possibility.

DISCUSSION

Overall we failed to replicate the significant findings reported earlier in an experiment assessing the presence of anomalous cognition in a situation closer to real-life. Two studies with 29 participants yielded chance results, and their effect sizes ($h = 0.04$ and $h = 0.11$) fail. For short of the effect size of $h = 0.42$ calculated from the data reported by Sheldrake & Smart [2]. However one of our participants differed significantly from these overall results. She passed the screening test with a significance of $p = .02$ and then scored constantly above chance in 8 series of 10 calls. Overall she reached a hit rate of 42.5% (MCE 25%). This corresponds to a p-value of .0002 and an effect size of $h = 0.37$. If all three studies presented here are combined, we arrive at 160 hits in 557 calls ($p = .02$, $h = 0.08$). This overall significance is attributable to this particular participant.

[2] This participant was a Bulgarian woman at that time visiting Freiburg, Germany. When the third author (VI) travelled to Sofia in summer 2007 she conducted the sessions for study 3 there in the participant's private home. As with all the experiments reported here, the sessions were conducted in German.

From our results it can be concluded that this kind of anomalous cognition, if it exists at all, is not commonplace as argued by Sheldrake & Smart [2]. Thus, it is also very likely that most of the anecdotal reports by people who claim in surveys to know who was on the phone can be accounted for by conventional explanations. Even the experimental findings of Sheldrake & Smart reflect this conclusion as they also selected their participants from a larger pool with screening tests, and their positive results were obtained with only 4 participants, with one of these four providing 68% of all trials. From this perspective both studies reveal that there are only a few selected participants who are able to score significantly above chance while the majority of participants show hit rates close to chance.

Overall, there were no significant differences between correct identification of familiar versus unfamiliar callers. But again, the story is different for our one successful participant: Her hit rate was 46% for familiar callers and 35% for unfamiliar callers. Although this difference is not significant in the χ^2-test, it is significant in a comparison of the *z*-scores (p=.04). A larger number of trials would be required to determine whether this tendency is real.

We cannot say with absolute certainty whether the positive results of study three represent a true effect, but we are confident that they do based on the fact that the they replicated the findings with this participant in study two. One additional cautionary note is that the same experimenter conducted all 80 trials, leaving open the possibility of the well-known "experimenter effect" often reported in parapsychological research [15-18].

Whether the higher ecological validity of study two is really necessary to demonstrate an alleged anomalous cognition effect could not be clearly answered. If the hypothesis holds true that only selected participants are able to score reliably above chance, then it might be just an accident that no such participant took part in study one.

Based on our findings we suggest that future research in this area should focus more on experiments with few well selected and gifted participants rather than testing the general population on their anomalous cognition ability. This has already be shown in other fields of anomalous cognition research [19]. If our data reflect a true effect then it may be necessary to invest larger efforts into the selection of participants. This could be either done on the basis of traits which were associated with anomalous cognition in earlier studies such as extraversion [20], belief in psi [21] or creativity [22-24], or as in our case by screening tests. We preferred the latter, more direct procedure since the reported individual differences show only small correlations with anomalous cognition. In contrast, screening tests similar to the final task directly assess the capacity under investigation. We developed an innovative adaptive design for study two which is suitable for this kind of research in a very flexible way without running into the problem of optional stopping. If participants are identified who are able to score repeatedly above chance, then we suggest varying experimenters systematically in order to study experimenters' influences on these results.

ACKNOWLEDGEMENTS

We would like to thank our participants. Stefan Schmidt and Harald Walach were supported by the Samueli Institute, Alexandria, VA at the time of conducting this study. Majella Horan and an anonymous reviewer helped us with the language of the manuscript.

REFERENCES

[1] Sheldrake R, Smart P. Experimental tests for the telephone telepathy. J Soc Psych Res 2003; 67: 184-99.

[2] Sheldrake R, Smart P. Videotaped experiments on telephone telepathy. J Parapsychol 2003; 67: 147-66.

[3] Brown DJ, Sheldrake R. The anticipation of telephone calls: a survey in California. J Parapsychol 2001; 65: 145-56.

[4] Sheldrake R. Telepathic telephone calls: two surveys. J Soc Psych Res 2000; 64: 224-32.

[5] Schmidt S, Müller S, Walach H. Letter to the editor. J Parapsychol 2003; 67: 409-11.

[6] Sheldrake R. Letter to the editor. J Parapsychol 2003; 67: 411-4.

[7] Lobach E, Bierman DJ. Who's calling at this hour? Local sidereal time and telephone telepathy. The Parapsychological Association, 47th Annual Convention, August, 5-8, 2004; Durham, NC: The Parapsychological Association 2004; pp. 91-7.

[8] Spottiswoode SJP. Apparent association between effect size in free response anomalous cognition experiments and local sideral time. J Sci Explor 1997; 11: 109-22.

[9] Müller S. Das telefonexperiment. Die experimentelle Umsetzung eines Alltagsphänomens, Albert-Ludwigs-Universtität Freiburg i. Br. 2003; [Unpublished Thesis].

[10] Marsaglia G, Zaman A. A new class of random number generators. Ann Appl Probab 1991; 1: 462-80.

[11] Burdick DS, Kelly EF. Statistical methods in parapsychological research. In: Wolman BB, Ed. Handbook of parapsychology. New York, NY: Van Nostrand Reinhold 1977; pp. 81-130.

[12] Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum 1988.

[13] Irwin HJ. An Introduction to parapsychology. McFarland: Jefferson, NC 1999.

[14] Erath D. Das Telefon-Telepathie-Experiment - Eine Replikationsstudie mit verbesserter ökologischer Qualität, Albert-Ludwigs-Universität Freiburg i. Br., Institut für Psychologie 2006; [Unpublished Thesis].

[15] Palmer J. confronting the experimenter effect. PSI-research methodology: a re-examination. Proceedings of an International Conference held in Chapel Hill, North Carolina October 29-30, 1988; 1988; New York: Parapsychological Foundation 1993; pp. 44-64.

[16] Smith MD. The role of the experimenter in parapsychological research. J Conscious Stud 2003; 10: 69-84.

[17] White RA. The Influence of experimenter motivation, attitudes, and methods of handling subjects on Psi test results. In: Wolman BB, Ed. Handbook of Parapsychology. New York: Van Nostrand, Reinhold 1977; pp. 273-301.

[18] Walach H, Schmidt S. Empirical evidence for a non-classical experimenter effect. An experimental, double-blind investigation of unconventional information transfer. J Sci Explor 1997; 11: 59-68.

[19] Utts JM. An assessment of the evidence for psychic functioning. J Sci Explor 1996; 10: 3-39.

[20] Honorton C, Ferrari DC, Bem DJ. Extraversion and ESP Performance: A meta-analysis and a new confirmation. J Parapsychol 1998; 62: 255-76.

[21] Lawrence TR. Gathering in the sheep and goats... a meta-analysis of forced-choice sheep-goat esp studies, 1947-1993. The Parapsychological Association 36th Annual Convention. Proceedings of Presented Papers; Durham, NC: The Parapsychological Association 1993; pp. 75-86.

[22] Dalton K. Exploring the links: creativity and psi in the ganzfeld. The Parapsychological Association 40th Annual Convention. Proceedings of Presented Papers; Durham, NC: The Parapsychological Association 1997; pp. 119-34.

[23] Holt N, Delanoy DL, Roe CA. Creativity, subjective paranormal experiences and altered states of consciousness. The Parapsychological Association, 47th Annual Convention, August, 5-8, 2004; Durham, NC: The Parapsychologic Association 2004; pp. 433-6.

[24] Bem DJ, Honorton C. Does PSI exist? Replicable evidence for an anomalous process of information transfer. Psychol Bull 1994; 115: 4-18.

Received: August 29, 2008 Revised: January 29, 2009 Accepted: February 18, 2009

© Schmidt *et al*.; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http: //creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.