

DO YOU KNOW WHO IS ON THE PHONE? REPLICATION OF AN EXPERIMENT ON TELEPHONE TELEPATHY

Stefan Schmidt, Susanne Müller, & Harald Walach

Department of Evaluation Research in Complementary Medicine

University Hospital Freiburg

ABSTRACT

Many people report that they know in advance who is on the phone when the telephone is ringing. Such reports may be explained by selective memory or expectancy effects but there are also examples that resist such hypotheses. Thus many people believe in an extrasensory communication transfer that may be termed 'telephone telepathy'. Surveys show that this kind of belief is widespread and might be one of the most common beliefs regarding the paranormal.

Rupert Sheldrake conducted several experiments to find out whether this effect is really due to ESP. Subjects had to determine which one of four possible callers is on the phone while the telephone was still ringing. Sheldrake reports highly significant hit rates that cannot be explained by conventional theories. He claims furthermore that callers who are familiar to the person answering the phone are identified at significant better rates than unfamiliar callers.

We attempted to replicate both of these findings by setting up a replication experiment. Twenty-one participants were twice invited for a two-hour session in an office-like room and were asked to pre-identify the callers of 10 phone calls during each session. The caller could be either one of two persons known to the participant or one of two persons unknown to them. With these four possible callers there is a mean chance expectation of 25% correct guesses. Participants were asked to complete questionnaires on earlier experiences with 'telephone telepathy', mood, personality and paranormal belief. Then participants spent approximately 100 minutes in the room together with an experimenter while the whole session was recorded on videotape. Approximately every ten minutes a telephone rang and the participant had to announce his call before the experimenter lifted the receiver to check who was calling. The sequence of the four possible callers was determined by a random event generator.

Overall valid responses were obtained on 397 calls and the participants identified 106 calls (26.7%) correctly. This result is not significant ($z = 0.78$). However, 67 (63.2 %) of the correct calls were by familiar callers showing that this group was identified more often than the unknown callers. But this result can be explained by a response bias in the participants because in all 397 calls they responded 242 times (61.0%) with the name of a familiar person. This resulted in a non-significant hit rate of 27.7% and also in a non-significant hit rate of 25.2 % ($z = 0.05$, n.s.) for the unfamiliar callers respectively. The difference between these two hit rates itself is not significant. Thus, this replication attempt failed to yield any telepathic effect. The possible reasons for the non-significant results are discussed.

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 person has an intuitive knowledge who is calling when s/he answers the phone. Or somebody is accidentally thinking about an old friend from school s/he hasn't heard from some twenty years just to find a message of this person on the answering machine the same evening. Friends call each other in exactly the same moment and both find the line engaged. Some of these 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 phenomena in California and the UK (Brown & Sheldrake, 2001; Sheldrake, 2000). 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." (Sheldrake & Smart, 2003a, p. 184). He was also the first person to conduct an experiment on this phenomenon.

In this simple experiment (Sheldrake & Smart, 2003a) a subject gets called by one of four pre-selected callers. The callers are selected by a random process and there is no conventional possibility to know in advance whom the caller will be. The subject guesses the name of the caller before answering the phone. Within this simple experiment there is a mean chance expectation (MCE) of $p = .25$ to make 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 exactly 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 unknown information transmission process (that might be called telepathy or just Psi), then the MCE should be exceeded in such an experiment, provided that the phenomena shows up on demand.

It is the merit of Rupert Sheldrake to develop such a simple and clear-cut design and to conduct this straightforward test for "telephone telepathy" as he calls it. In his first publication, he and his colleague report on several series of altogether 571 trials with 63 different participants (Sheldrake & Smart, 2003a). 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 telepathy hypothesis was apparently confirmed by an extraordinarily strong effect. However, on close examination this study showed some 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 participant and asked what s/he had guessed. In some trials this was 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 more serious. When the telephone rang the participants lifted first the receiver and then spoke aloud the name of the guessed person 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.

But a second study (Sheldrake & Smart, 2003b) with four participants was controlled much tighter. For this study only selected participants were employed. Every 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 as a valid and good controlled examination of the hypothesis on telephone telepathy. Overall there were 122 correct guesses (hit rate 45%, $p = 10^{-12}$) and the authors conclude this to be a proof for telepathy. They also tested a second hypothesis on the difference between callers who are known to the participants and others 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 towards the familiar callers there will be a telepathic bond with 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 a significant difference ($p = 8 \times 10^{-8}$). However the statistical procedure they applied proved to be unsatisfactory. The data were not corrected for a response bias. Upon inspection of the data it could be shown that the participants indeed guessed more often familiar callers than unfamiliar ones and this resulted in a higher probability to obtain a hit with a familiar caller (Schmidt, Müller & Walach, 2003). When corrected for this bias the hit rates resulted in 46% hit rate for familiar callers and 37% for unfamiliar callers respectively. The difference between these hit rates was no longer significant ($p = .32$), if

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

the same statistical test that was also used by Sheldrake & Smart (2003b) was applied. However, in a reply to this critique Rupert Sheldrake showed that the data are reaching significance again if a different statistical analysis is applied (randomization test, $p=0.0002$). With the data being significant in one analysis and not significant in the other, the question whether familiar callers are identified better than unfamiliar ones remains open.

Altogether the Sheldrake studies have a rather exploratory nature with frequent changes in the design and only partially prespecified number of trials. We thus set out to conduct a more rigorous conceptual replication of these two findings (telephone telepathy effect, better results with familiar callers). In this experiment we repeated the basic design of the Sheldrake studies. We invited participants who report to be familiar with the phenomenon and asked them to name two callers. Two other callers were assigned by the experimenter and thus the participant had the task to identify the correct caller out of a possible four of which only two were familiar to him or her. However, to improve on experimental controls and to make the data collection more efficient we also applied changes to the original protocol. Instead of visiting the participants at home we invited them to an office. This provided us with the possibility to conduct all trials under the same controlled conditions, but on the other hand also reduced the ecological validity of the study. We furthermore carried out ten consecutive trials at once within a two-hour session. We also decided to work with a self-selected sample rather than to identify high-scorers by pre-tests. This approach was chosen to guarantee that the results generalize to the reports from everyday life. Another change applied to the role of the experimenter. In contrast to the Sheldrake studies, where there was no experimenter present and where the participants answered the call, the experimenter in our study stayed with the participant all the time during the experiment. The experimenter also lifted the receiver for the incoming calls after the participant's guess was recorded to find out who was calling. We did so to avoid a sensory leakage in a possible participant-caller interaction. This was necessary because callers knew whether they had to call again in the same session or not and thus might drop accidentally verbal cues such as "till later" or "OK. We are done!"

METHODS

Design and Hypotheses

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. Hypothesis 1 stated that participants are able to identify the correct caller out of a possible four significantly better than expected by chance (MCE = 25%). Hypothesis 2 stated that participants are able to identify familiar callers significantly better than unfamiliar ones.

Participants

Twenty-one participants were recruited through newspaper advertisement, press reports and leaflets. Nineteen participants contributed 20 trials in two sessions of ten. Two participants took part in only one session of ten trials. Seventeen participants (81%) were female and the mean age of all participants was 39.8 years (SD = 11.3, range 22-62).

Experimenters

Three different experimenters participated in the study. One of them, a 35 year-old psychologist served as the experimenter for the interaction with the participants for all sessions. The other two (the second author SM and a student) were in charge of scheduling the callers during the session and had no direct interaction with the participants.

The complete organization and conductance of this study was carried out by the second author, SM (Müller, 2003).

Facilities

The experiment took place in an office space of a youth counselling organisation in Freiburg, Germany. The office contained a video camera for the monitoring of the session. The telephone for the incoming calls was placed on a table. The telephone had no display for caller identification. The room was further equipped with reading material for the participants and some cookies.

For the randomization and scheduling of the callers a second office was used in the University Hospital Freiburg. This office was located in a distance of approximately 3 km from the first one.

Material

Participants were asked to complete the following questionnaires:

1. A questionnaire similar to the one presented by Shelldrake (2000) asking about prior experiences with 'telephone telepathy'. In addition we added some questions on the expectancy of success in the upcoming experiment.
2. A German sheep-goat scale. We applied the questionnaire on paranormal beliefs by Brednich (1993), that was already applied in other German Psi studies (Schmidt, Schneider, Binder, Bürkle & Walach, 2001; Schmidt, Tippenhauer & Walach, 2001; Walach & Schmidt, 1996). This is 21-item scale with various statements on ESP, reincarnation, magic and astrology that have to be rated on a four-point scale.
3. The German translation of the NEO-FFI (Borkenau & Ostendorf, 1993) a sixty-item personality inventory assessing the five standard traits: neuroticism, extraversion, openness, conscientiousness and agreeableness.
4. The Basler Mood Scale/Basler Befindlichkeitsskala, (Hobi, 1985). A 16-item scale assessing the present mood of the participant on the basis of a subjective rating. The scale has four factors: vitality, balance, alertness and social extraversion.

Procedures

Participants interested in the advertised study called the experimenter SM and were asked whether they already had experiences of knowing who was on the phone in the absence of sensory cues and whether they would be able to provide two callers to participate in the experiment. If these criteria were fulfilled, two sessions were scheduled. Upon arrival at the location the participants were greeted by the experimenter and informed about the procedures of the session. Then the participant was asked to fill in the questionnaires. The first three questionnaires were only filled in once during the first visit. The Basler Mood Scale was filled in prior to both experimental trials.

While the participant filled in the questionnaires, the experimenter staying with the participant switched on the video camera and then phoned the second experimenter in the second office at the hospital. The second experimenter then collected a sealed envelope from a locked cabinet drawer that contained the randomized calling sequence for the session. The sequence consisted of the numbers 1, 2, 3, and 4, and were assigned as follows to the four callers: 1 to the first familiar caller that was nominated by the subject, 2 to the second familiar caller, 3 to the first unfamiliar caller chosen by the experimenter and 4 to the second one. The ten calls were then scheduled 10 minutes apart. The scheduling experimenter served as one of the unfamiliar callers. The other three callers were then called by the experimenter and informed about the telephone number to call and the calling time. If a caller was chosen by the randomization to conduct more than one call then the schedule for all calls was given at once. Once the scheduling was completed, the

scheduling experimenter phoned back to the experimenter staying with the participant and the experiment was ready to start. During the next 90 minutes the participant received ten calls approximately one every ten minutes. When the telephone rang the participant spoke his or her guess into the video camera and the experimenter recorded the guess on a session sheet. Then the experimenter lifted the receiver and checked for the caller. This information was fed back to the participant. Between the calls the participants were free to read some of the magazines provided, whose content was unrelated to the experiment, or they could have a chat with the experimenter.

When the participant arrived for the second sessions the procedures were the same except that the participants this time only had to fill in the mood scale but not the other questionnaires.

Randomization, Data analysis and Statistics

For the randomization, the Marsaglia-Zaman random number generator was employed (Marsaglia & Zaman, 1987). This RNG consists of a software algorithm that deduces two seeds from the computer clock and produces random sequences that have proven to be perfectly at chance. Fifty sequences of ten number sequences consisting of the numbers from one to four were generated before the first session of the experiment and placed in opaque envelopes labeled with a number from 1-50. These envelopes were stored in a locked cabinet drawer. The experimenter always chose the envelope labeled with the smallest number.

Questionnaire data, calling sequences and guesses of the participants were entered into SPSS for Windows 11.0. Further analyses were conducted in SPSS and in Excel.

Statistics: The mean chance expectation for each single trial is $p = .25$ and within each trial all four targets have the same probability independently of the prior trials (open deck). For Hypothesis 1, the likelihood of X hits within N trials can be calculated by the classical formula for the critical ratio (CR): $CR = (X - Np)/\sqrt{Npq}$ (with $q=1-p$) (Burdick & Kelly, 1977). 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 the prespecified method for the evaluation of hypothesis 1. The test of significance was one-tailed. Another possibility to assess the results is the application of a binominal test calculating the likelihood to obtain X or lesser hits out of N trials, when the probability to obtain a hit is at $p = .25$.

For Hypothesis 2 we calculated a χ^2 -test for the distribution of the data in 2×2 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 from the 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. The calls came from the following callers: familiar caller 1 (f1): 103 calls; familiar caller 2 (f2): 98 calls; unfamiliar caller 1 (u1): 101 calls; unfamiliar caller 2 (u2): 95. These numbers prove that the random distribution was indeed at random.

On the other hand, the distribution of the guesses for these four callers by the participants was not at random. Participants guessed in most cases the caller to be a familiar one with an additional preference for the first familiar caller they had named. The guesses were: guessed for f1: 131; guesses for f2: 111, guesses for u1: 96; guesses for u2: 59. This distribution shows a response bias.

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 $z = .78$. The according p -value (one tailed) is $p = .22$. The binominal test arrives at a similar p -value of $p = .23$. Thus, hypothesis 1 was not confirmed.

Out of 19 participants that completed all 20 trials two arrived at an individual significant hit rate of 9 and 10 hits respectively. The two participants completing only ten trials had a non-significant score of two hits each.

Hypothesis 2: From the 106 correct guesses 67 (63.2%) were with familiar callers and 39 (36.8%) with unfamiliar callers. If based on the distribution of the incoming calls this would result in a hit rate of 33.3 % for familiar callers and 19.9 % for unfamiliar caller. In other words when a familiar caller called s/he was identified in 33.3% of all trials correctly as familiar. But it was already said in the introduction that such a procedure to base the hit rates on the distribution of the incoming calls is prone to error because it does not take into account the response bias of the participants. If the hit rates are based on the participants' guesses for familiar or unfamiliar callers the hit rates change to the following values: familiar callers: 27.7 % (n.s.) and unfamiliar callers: 25.2 % (n.s.). The according 2 x2 table looks as follows (see Table 1).

		Hit		Sum
		Yes	No	
Guesses	Familiar	67	175	242
	Unfamiliar	39	116	155
	Sum	106	291	397

Table 1: Crosstabulation of hits and guesses for familiar and unfamiliar callers.

The χ^2 -test for this table results in a non-significant p-value of $p = .58$ ($\chi^2 = 0.31$, $df = 1$) and thus hypothesis two was not confirmed.

For the sake of completeness we provide the raw data of all 21 subjects in Table 2.

Part. No.	Trials	Guess Fam.	Guess Unfam.	Hits Fam.	Hits Unfam.	Total Hits	Total Hits %
1	20	10	10	2	3	5	20
2	19	11	8	3	2	5	26
3	19	15	4	3	2	5	26
4	19	12	7	4	2	6	32
5	20	10	10	3	0	3	15
6	10	6	4	1	1	2	20
7	20	8	12	1	3	4	20
8	20	13	7	6	2	8	40
9	10	7	3	1	1	2	20
10	20	12	8	4	0	4	20
11	20	12	8	4	3	7	35
12	20	14	6	5	1	6	30
13	20	14	6	2	1	3	15
14	20	13	7	1	0	1	05
15	20	11	9	3	1	4	20
16	20	14	6	3	2	5	25
17	20	13	7	6	3	9	45
18	20	13	7	7	3	10	50
19	20	16	4	2	1	3	15
20	20	8	12	4	4	8	40
21	20	10	10	2	4	6	30
total	397	242	155	67	39	106	

Table 2: Complete data for all 21 subjects and 397 valid trials. guess fam. = number of guesses for familiar callers; guess unfam. = number of guesses for unfamiliar callers; hits fam. = number of hits for incoming calls of familiar callers; hits unfam. = number of hits for incoming calls of unfamiliar callers.

DISCUSSION

Our task to replicate the positive findings on telephone telepathy reported by Rupert Sheldrake in a tighter controlled setting failed. There was neither a hit rate larger than expected under chance condition, nor did we find any significant difference between the hit rates for familiar and unfamiliar callers. The data obtained in our study match almost perfectly chance expectation and do not allow for any interpretation of unconventional information transfer.

This result is in accordance with the hypothesis that all accounts for strange experiences as described in the introduction are due to coincidences, selective memory or unconscious expectancy and the like. But this result is definitive not in accordance with the findings reported by Sheldrake and Smart (2003a, 2003b).

There are various interpretations to this inconsistency. One would be that there is no such effect like telephone telepathy and that the data presented by Sheldrake are due to some methodological error. The second one would be that one of the changes to the design that were made by us in our undertaking to establish a conceptual replication was fatal and avoided the effect demonstrated by Sheldrake to show up in our study. A third interpretation would centre on the replication problem within parapsychology and would speculate about basic properties of Psi that avoid replication beyond ones that can be attributed to changes in the design.

The first interpretation that the Sheldrake findings are based on erroneous methods is of course one possibility. But one has to acknowledge that the experimental procedures described by Sheldrake and Smart (2003a, 2003b) show a steady increase in controls and thoroughness from the first pilot trials to the final design with videotaped sessions and significant effects are reported for all of these different protocols. Of course one can never be sure what really happens in these studies until one witnesses the data collection, but from the information that was available to us these data look valid and cannot be explained away.

Thus, the second interpretation stating that differences in the design between our study and the ones by Sheldrake are responsible for our failure to replicate Sheldrake's result is becoming more likely. The main changes in our experiment were: (i) The experiment took place in a room provided by us rather than in the participants' homes. (ii) The participants received ten calls and not one per session. (iii) The experimenter was present during the session and also lifted the receiver instead of the participants. (iv) Selection of sample: we used a self-selected sample rather than to restrict the experiment to high scorers who have passed a pretest.

The first three changes reduced the ecological validity of the study (in favour for better controls) and thus one could conclude that the phenomenon shows up under natural conditions (participants at home by themselves getting one phone call) but not under the more artificial conditions in the setting we provided (participant receiving phone calls every ten minutes in an office that is alien to them). On the other hand, one could argue that Sheldrake also obtained significant findings in sessions that were placed and broadcasted from a TV studio. In Sheldrake's earlier trials participants also were called twice and the performance did not decrease (although one has to mention that there is some likelihood for sensory cueing in these trials).

Whether the differences introduced by the changes (i) and (iii) are really responsible for our failure to replicate can be only checked through a direct comparison of these conditions in a follow-up study. But for the change in the number of successive calls (ii), we can check whether participants were more successful in their first trials. Figure 1 shows the hit rate per trial (bars) and also a curve representing the cumulative hit rates for the trials 1 to 10 of a session. From the cumulative curve one can extract the results if one would stop after a shorter number of trials than 10 in one session. For only one trial would one obtain the highest hit rate (30.8%, $p = .25$). If our participants had shown the same hit rate in all 397 trials as in their first trials, we would arrive at a significant p-value of $p = .006$. Thus it may also be that the difference between our results and the results by Rupert Sheldrake are at least partially due to the fact that participants performance decreased after their first trial. One reasonable explanation would be that participants get tired over time. On the other hand, one can see that this might also be due to a normal fluctuation. An

inspection of the average hit rate in the different trials (bars) demonstrates that participants performed even better in their 6th and also in their 9th trial than in their first one.

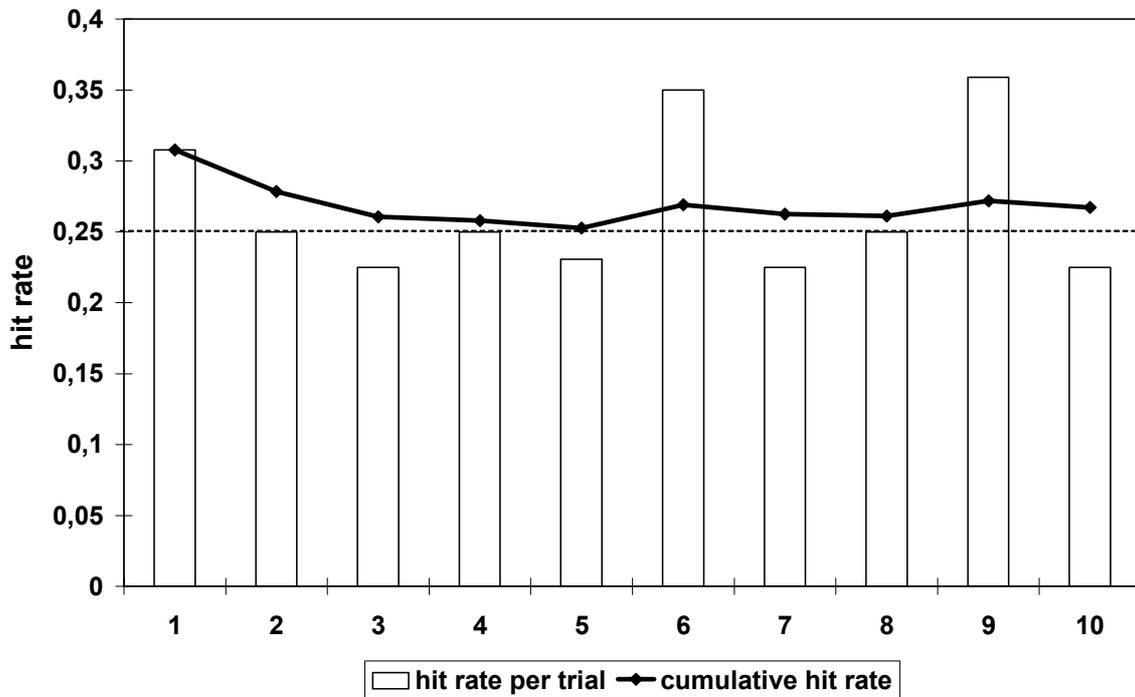


Figure 1: Hit rate per trial (bars) and cumulative hit rate (curve) over the successive 10 trials in a session. The dotted line indicates the MCE of .25.

The change in location and the presence of the experimenter also might have been changes that mattered. One of the safeguards in our study was that the experimenter, rather than the participant (and thus receiver of the call), lifted the receiver and talked to the callers. One could argue that by this procedure the effect was avoided because the caller did not have the chance anymore to reach the participant. But on the other side one could argue as well that the meaning context of the experiment was still intact. The participant sat in front of the telephone and tried to find out who is responsible for the ringing of the phone. The change in the procedure occurred after the recording of the guess and furthermore everybody knew about this circumstance. The point in this discussion is that to find out which of these two lines of reasoning is the right one, one has to know the precise mechanism of the alleged effect. Until this mechanism is known one can not give a precise experimental description how to find the effect and also does not know what effect size under which conditions can be expected. Collins calls this paradox the experimenter regress (Collins, 1985).

An important difference was also the selection of participants (iv). One could argue that Sheldrake has selected sensitive participants through his process of performing a pre-test. Other researchers have already reported that telepathy is not necessarily normally distributed in the population but restricted to a few gifted subjects (see e.g. Utts, 1996). Our self-selected sample probably did not contain any high scoring participant (or maybe not enough). But if this were true then the many reports of strange experiences around receiving phone calls could not anymore be attributed to 'telephone telepathy'. If this is really "... the commonest kind of apparent telepathy in the modern world" that is reported by approx. 50% of the population (see Introduction) then the experiment should also work with a self-selected sample that has experience with this phenomenon. It was our specific aim to study this everyday life situation and we recruited our sample accordingly.

Finally there remains the possibility that our result is due to an intrinsic property of Psi not to show up stable within the same experimental context. Not to find the same effect again in a replication is a well-known result within parapsychology and has also occurred in our work before (see e.g. Schmidt, Tippenhauer & Walach, 2001). A theoretical framework for this can be found, for example, *within Lucadou's Model of Pragmatic Information (Lucadou, 1995, 2001). Whether such a circumstance is responsible for our results can be only determined by the development and empirical testing of a model for Psi that accounts for such intrinsic properties.

CONCLUSION

The positive findings by Sheldrake and Smart (2003a, b) could not be replicated in our experiment. This may be due to changes that were made to the design due to our task to conduct a conceptual replication with tighter controls. Whether these changes were responsible for our failure to replicate earlier findings cannot be clearly decided. This can only be determined by a replication where these conditions can be compared (e.g. the results of the same participants are compared in the experimenter's office and in their home, or the results of the same participants are compared in a condition with an experimenter present or not present). We suggest conducting further studies under various conditions by different experimenters and researchers to shed light on conducive and inhibitory conditions of the alleged effect.

REFERENCES

- Borkenau, P. & Ostendorf, F. (1993). *Neo-Fünf-Faktoren Inventar*. Göttingen: Hogrefe.
- Brednich, A. (1993). Eine Fragebogenuntersuchung zur Erfassung von magisch-irrationalen Denken und schizotypischen Persönlichkeitsstörungen bei Erwachsenen. Unveröffentlichte Diplomarbeit, Albert-Ludwigs-Universität Freiburg i. Br.
- Brown, D. J., & Sheldrake, R. (2001). The anticipation of telephone calls: A survey in California. *Journal of Parapsychology*, 65, 145-156.
- Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In: B. B. Wolman *Handbook of parapsychology* (pp. 81-130). New York, NY: Van Nostrand Reinhold.
- Collins, H. M. (1985). *Changing order*. London: Sage.
- Hobi, V. (1985). Basler Befindlichkeits-Skala. Ein Self-Rating zur Verlaufsmessung der Befindlichkeit. Weinheim: Beltz Test Gesellschaft.
- Lucadou, W. v. (1995). The model of pragmatic information (MPI). *European Journal of Parapsychology*, 11, 58-75.
- Lucadou, W. v. (2001). Hans In luck: The currency of evidence in parapsychology. *The Journal of Parapsychology*, 65, 3-16.
- Marsaglia, G., & Zaman, A. (1987). Toward a universal random number generator (No. FSU-SCRI-87-50): Florida State University.
- Müller, S. (2003). *Das Telefonexperiment. Die experimentelle Umsetzung eines Alltagsphänomens*. Unpublished thesis, Albert-Ludwigs-Universität Freiburg i. Br.
- Schmidt, S., Müller, S., & Walach, H. (2003). Letter to the editor. *Journal of Parapsychology*, 67, 409-411.
- Schmidt, S., Schneider, R., Binder, M., Bürkle, D., & Walach, H. (2001). Investigating methodological issues in EDA-DMILS: Results from a pilot study. *Journal of Parapsychology*, 65, 59-82.
- Schmidt, S., Tippenhauer, G., & Walach, H. (2001). Connectedness between participants in an experiment on distant intention. *The Parapsychological Association 44th Annual Convention. Proceedings of Presented Papers* (pp. 285-300). Durham, NC: The Parapsychological Association.

- Sheldrake, R. (2000). Telepathic telephone calls: Two surveys. *Journal of the Society for Psychical Research*, 64, 224-232.
- Sheldrake, R., & Smart, P. (2003a). Experimental tests for the telephone telepathy. *Journal of the Society for Psychical Research*, 67, 184-199.
- Sheldrake, R., & Smart, P. (2003b). Videotaped experiments on telephone telepathy. *Journal of Parapsychology*, 147-166.
- Utts, J. M. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration*, 10, 3-39.
- Walach, H., & Schmidt, S. (1996). Evidence for a non-classical experimenter effect. (Abstract). *Society for Scientific Exploration Third European Meeting*.

Address for correspondence: Stefan Schmidt, University Hospital Freiburg, Germany, Department of Evaluation Research in Complementary Medicine, Institute of Environmental Medicine and Hospital Epidemiology, Hugstetter Str. 55, D-79106 Freiburg, Germany.
E-mail: stefan.schmidt@uniklinik-freiburg.de