Automated Tests for Telephone Telepathy Using Mobile Phones

Rupert Sheldrake, PhD# Pamela Smart, and Leonidas Avraamides

Objective: To carry out automated experiments on mobile phones to test for telepathy in connection with telephone calls.

Study Method: Subjects, aged from 10 to 83, registered online with the names and mobile telephone numbers of three or two senders. A computer selected a sender at random, and asked him to call the subject via the computer. The computer then asked the subject to guess the caller's name, and connected the caller and the subject after receiving the guess. A test consisted of six trials.

Interactions Evaluated: The effects of subjects' sex and age and the effects of time delays on guesses.

Main Outcome Measure: The proportion of correct guesses of the caller's name, compared with the 33.3% or 50% mean chance expectations.

Main Results: In 2080 trials with three callers there were 869 hits (41.8%), above the 33.3% chance level ($P < 1 \times 10^{-15}$). The hit rate in incomplete tests was 43.8% ($P = .00003$) showing that optional stopping could not explain the positive results. In 745 trials with two callers, there were 411 hits (55.2%), above the 50% chance level ($P = .003$). An analysis of the data made it very unlikely that cheating could explain the positive results. These experiments showed that automated tests for telephone telepathy can be carried out using mobile phones.

Key words: telepathy, ESP, automated tests, telephones

(Explore 2015; 11:310-319 © 2015 Elsevier Inc. All rights reserved.)

Most people have had seemingly telepathic experiences in connection with telephone calls. They think of someone for no apparent reason, and then receive a call, or else they know who is calling when they hear the phone ring before they answer their telephone or look at the caller identification display. Many people have had similar experiences in connection with mails and text messages.[1] Are these experiences just a matter of chance coincidence, selective memory, or wishful thinking?

There have now been many experiments to test where people really can tell who is calling them or sending them a message. In these previous tests, the subjects nominated three or four people they knew well, and these people served as contacts for the experiment. During the tests, the subjects and contacts were in different locations, often many miles, sometimes thousands of miles, apart. For each trial, one of the contacts was selected at random and asked to call or send a message to the subject. Before receiving the call or message, the subjects were asked to guess which of these contacts was getting in touch with them. Some of these tests were carried out under the direct control of experimenters, and others were automated. Some were carried out under loosely controlled conditions, and in others the subjects were filmed continuously to ensure that they were not receiving messages through other telephones or means of communication that could have enabled them to cheat. The results of these experiments are summarized in Table 1. All of them gave above-chance results.

For technical reasons, it was easier to set up automated tests with emails and text messages than with telephones, but in 2008, automated telephone telepathy tests became echnically feasible for the first time. In this paper, we describe the results of two such experiments carried out with mobile phones. The tests in the first experiment involved four participants, one subject and three contacts. In each trial the subject had to guess which of the three contacts was calling, and hence the chance hit rate was 33.3%. In the second experiment, there were three participants, one subject and two contacts, and hence a chance hit rate of 50%.

In this study, our primary aim was to explore the feasibility of large-scale experiments under real-life conditions. The rapid technical advances in telephone systems make possible new applications for testing telepathy and other aspects of human psychology. This paper is a preliminary exploration of the research possibilities opened up by these new technologies. In particular, we were interested in finding out how easy it is to recruit participants, and what patterns of results we would obtain under "ecological" test conditions, in which participants were using mobile phones while carrying on with their normal lives.

We did not film or supervise the participants, and hence it was possible that some were cheating. Therefore, we do not claim that positive results in these exploratory experiments are compelling evidence for telepathy, but rather that they point to new ways of investigating telepathy experimentally, and

Corresponding author: 20 Willow Rd, London NW3 1TJ, UK.
e-mail: rsheldrake@clara.co.uk

ISSN 1550-8307/$36.00

http://dx.doi.org/10.1016/j.explore.2015.04.001

Table 1. Summary of Previous Research on Telepathy in Connection With Telephone Calls, Emails, and SMS Messages

Experiment	Trial Number	Chance (%)	Hits (%)	*P*	Reference
Telephone tests					
Unfilmed	571	25	40	1×10^{-15}	2
Filmed	271	25	45	1×10^{-12}	3
Filmed	12	25	50	.05	4
Filmed	216	25	29.4	.05	5
Filmed	557	25	28.7	.02	6
Email tests					
Unfilmed	552	25	43	1×10^{-18}	7
Filmed	137	25	47	1×10^{-7}	7
Automated	419	33	41.8	.0001	8
SMS tests					
Automated	886	33	37.9	.001	9

suggest how automated techniques can be improved to overcome problems revealed by this study.

A secondary purpose of our experiments was to investigate if the sex and age of the subjects had any noticeable effects on hit rates. In transnational surveys, more women than men reported telepathic experiences,[10,11] and more women than men claimed to have experienced apparent telephone telepathy.[1] Some surveys showed a general tendency for claims to telepathic experiences to increase with age.[11] However, in previous automated telepathy tests in connection with emails[8] and short message service (SMS) messages,[9] there was no significant effect of sex or age. We therefore wanted to explore the effects of sex and age in these automated telephone telepathy tests, but had no particular expectations or hypotheses.

METHODS

Procedure

This test was designed to be easy to carry out by subjects and their contacts on mobile telephones anywhere where they had reception. The only restriction was that the subjects should be in different locations from their contacts, and that should not communicate with each other during the trials by emails or landline telephones, or any other means.

Participants registered online through Rupert Sheldrake's (R.S.) web site, www.sheldrake.org. The subjects entered their first and second names, sex, age, mobile telephone number, and email address, and also entered the names of two or three contacts (first names only) together with their mobile telephone numbers. The subject was told, "During the test, when you receive a call you will be asked to guess whether it is from contact 1, 2, or 3 (or 1 and 2 in the case of the two-caller test) so you will need to remember the order of your contacts. It will help if you put them in alphabetical order."

There was also a field on the registration form for a group name, so when participants were part of a specific group, they all entered the same group name when registering, enabling their data to be retrieved as a group. The subject then received a welcome SMS message saying, "Thank you for entering the Telepathy Test which will start shortly. Your PIN is <nnnn>. Good luck!" The personal identification number (PIN) was a four-digit number, specific to this test. The contacts also received an SMS message saying, "Your details have been submitted by <SubjectName> as part of the Telepathy Test and the test will start shortly. Your PIN will be <nnnn>." [The subject was also told that she could stop the test at any time by calling the (landline) telephone telepathy test number (which was given at the bottom of the registration form) and pressing the star key on the keypad.]

Thus all participants' tests were pre-registered, and hence there were no data from this test in "file drawers." The test proceeded as follows:

1. After a random time delay of between 1 and 10 min, the system selected one of the contacts at random and sent a message saying, "This is the Telepathy Test. Please call <landline number> to be transferred to <SubjectName>. Your PIN is <nnnn>. Do not attempt to contact <SubjectName> directly."
2. The contact person then called the telephone telepathy test landline number and was asked to enter the PIN number, identifying which test the contact was part of. A voice message asked the caller, "Please stay on the line while we attempt to contact the subject." While on hold, the subject heard music.
3. The computer then telephoned the subject, whose caller ID display said, "Telephone telepathy test." When the subject answered the phone, a message said, " One of your callers is on the line. Please guess who it is by pressing 1, 2 or 3 [or 1 or 2 in two callers tests]." As soon as the guess was made it was recorded automatically, and the line opened up so the caller could talk to the subject, thus receiving immediate feedback. After a minute, the call was terminated.
4. After a random time delay of between 1 and 10 min, this procedure was repeated, and then repeated again until the subject had completed six trials, at which stage the test was complete. The subject then received an SMS message

saying "Thank you for taking part in the Telepathy test. You scored <CorrectAnswer> correct out of 6 trials." The contacts also received SMS messages saying, "Thank you for taking part in the Telepathy test. Subject scored <CorrectAnswer> correct out of 6 trials."

The data were stored on an online database, accessible by the use of a password. The database displayed a chronological list of results, with one line per test, giving the subject's name, the subject's sex and age, the date and time at which the test started, the number of trials, and the number of hits. Also, separate columns showed the number of trials and hits that occurred when contacts called the subject within four minutes of receiving a request to do so, those that occurred 4–10 min afterwards, and those that occurred more than 10 min afterwards. In addition, for each test it was possible to display all the details trial by trial, with a full list of all messages sent and received by the system during the test, together with the time at which they were sent or received, recorded to the nearest second. The database could be searched by group name, subject's name, subjects' ages, and sex, and data from complete tests could be displayed separately from data from incomplete tests, in which there were fewer than six trials.

System and Programming

The system was developed and operated by Mobifi Ltd, a London-based SMS solutions and application provider. Mobifi had an email gateway that enabled customers to send and receive email messages. The telephone calls were sent and received through a platform operated by X-on (Storacall Ltd). X-on was a provider of Interactive Voice Response (IVR) services.

The core of the application was written in Transact-structured query language (T-SQL), the language for operations with databases in SQL server. The version used for this application was SQL Server 2003. Microsoft SQL had its own developer utilities, which were sufficient for developing this application. The standard Microsoft SQL procedures for generating seed numbers and random numbers were used to select the contacts for each test. The random numbers were thus derived from a standard random-number algorithm, rather than from a random-event generator.

Parts of the application were programmed in Visual Basic Script (VB Script). This script was used to present the results on the web. All the web pages were ASP pages, a version of VB Script.

The web part of the application was used to create the reports and test results, using Microsoft Internet Information Server (part of the Windows Server 2000 operating system). All data about subjects, contacts, registration details, test activity, and test results were held in the SQL Server, as were all procedures for evaluating message delivery and information about calls from X-on services.

All procedures for sending SMS to participants and sending commands to X-on were developed in the web part of the application. The process of generating random selections was created using RAND (http://msdn.microsoft.com/en-us/library/ms177610.aspx), a built-in function of Microsoft SQL Server, which uses a pseudo-random float value, dependent on a seed value given by the exact timing of the random-number request. The application was hosted on a dedicated server that was located in a secure data center in Byfleet, Surrey, UK. The operating system was a Microsoft Server 2000, which was managed and maintained by Mobifi Ltd. The server could be accessed only by the server administrators, namely Mobifi staff and the developer of the application. This server and the web application had separate passwords and were therefore protected. Data could be entered to the SQL database only by the application and the relevant hypertext transfer protocol (HTTP) posts (results of calls to participants of a test) by X-on, a provider of IVR services.

The application sent commands to the IVR platform via HTTP to initiate calls to participants of tests. Once the outcome of the calls has been determined, X-on sent a HTTP to the application. The groups and user data were retained on the X-on service unless removal was requested by Mobifi. This allowed for an upper limit of 10,000 subjects. The user data were kept securely on a Microsoft SQL Server as part of X-on's network, and user data were never modified by X-on unless specifically requested by Mobifi.

The IVR services operated using Envox/Dialogic software and were controlled by the database tasks, and the database data were controlled by the .NET web service pages used by Mobifi to register and initiate tests. These were all under X-on's control and were configured to transfer data reliably and securely. At no time was the setup modified or accessible to technicians not involved in the project. All test initiation and subject/contact selection were controlled by Mobifi and the X-on system was configured to respond to requests only from the Mobifi servers.

The X-on IVR service received and sent calls out over the public switched telephone network (PSTN) in the same way in which home phones generally connect to each other. It used a single carrier to deliver the calls and it was not possible for calls to be altered when traveling over the network. Effectively, the IVR service received and sent calls like a typical phone, apart from the fact that the IVR made and received all calls rather than someone physically dialing.

Participants

Participants were recruited in two ways. First through R.S.'s web site (www.sheldrake.org), where anyone could volunteer to take part. Second, by Research Helpers, volunteers who were taking part under the aegis of the Perrott–Warrick Project, of which R.S. was the Director. These were people who wished to gain some research experience. Each of them was asked to do the test themselves and then recruit other subjects from among their friends and family members. Rupert Sheldrake's Research Assistant, Pam Smart, guided them by explaining the procedure, and answered any queries by telephone or email. The Research Helpers themselves usually served as a contact in the tests with the participants they recruited, and usually registered the subjects. Research Helpers were paid a modest sum (£100) for doing the test themselves and recruiting six other subjects, and writing a short report about their group's experiences with the test. Subjects were between 10- and 83-year old, with the majority aged between 20 and 29. All participants were told that they should do tests only when the subjects and all contacts were free to make telephone calls and not, for example, when any of them was on plane journeys or in classes or meetings. They were also told that contacts should not be in the same place as those of subjects and should not communicate with them by other means during the tests.

It was not easy to recruit participants for these tests, because of low motivation and because subjects and/or their contacts were often busy and found it hard to agree on a time when they would all be free to participate in the test.

We originally planned to conduct 2000 trials with three callers, and when this had been achieved we repeated the experiment with two callers, to find out whether this made it easier to recruit participants, since each group required only three people rather than four. Recruitment was indeed somewhat easier, and in this follow-up study we aimed to carry out at least 700 trials. In fact we completed 745.

Statistical Analysis

The results were analyzed using the binomial test, by pooling together the results from all subjects. The chance probability of a hit was 0.33 in the tests with three contacts and 0.5 in the tests with two contacts. Single-sided P values were used because all previous research on telephone, email, and SMS telepathy had shown positive results (Table 1), and our hypothesis was that in this automated test we would also see positive, above-chance results. The null hypothesis was that the hit rate in the tests would not be significantly different from that of the chance level. For comparisons of male and female subjects, the 2×2 chi-squared test was used. For analysis of the effects of subjects' ages, age was modeled as a categorical variable with five levels using a generalized linear model (GLM), with quasibinomial errors and a logit link function. Cohen's effect size d was calculated for tests with three callers according to the formula $d = [P(\text{hits observed}) - 0.33]/\sqrt{(0.33 \times 0.66)}$, where P (hits observed) is the proportion of hits, and for tests with two callers according to the formula $d = [P(\text{hits observed}) - 0.5]/\sqrt{(0.5 \times 0.5)}$.

RESULTS

Overall Results: Tests With Three Callers

In tests with three callers, there were 2080 trials altogether, with a hit rate of 41.8%, well above the 33.3% chance level ($P < 1 \times 10^{-15}$). The effect size (Cohen's d) was 0.19.

There were 288 complete six-trial tests, with 1728 trials, in which the hit rate was 41.6% ($P < 1 \times 10^{-12}$). The 95% confidence interval for this figure was from 39.2% to 43.9%. There were 349 incomplete tests, in which the hit rate was 154 out of 353 trials (43.8%; $P = .00003$) (Table 2). This shows that the positive hit rates in completed tests could not be explained by "optional stopping," whereby people who are not scoring well give up before completing the test. If optional stopping were important, then the hit rate in incomplete tests would be lower than in complete tests, but in fact it was higher. An analysis of the detailed trial-by-trial data showed that most incomplete tests came to a stop not because the subject gave up but because one of the callers failed to respond to a request to call the subject. In addition, 264 tests were registered for which there were no trials, and hence no hits.

Table 2. Telephone Telepathy Tests With Three Callers: Hit Rates in All Tests, Complete Tests and Incomplete Tests

Tests	Hits	Trials	Hits (%)	P
All	869	2080	41.8	$<1 \times 10^{-15}$
Complete	718	1728	41.6	$<1 \times 10^{-12}$
Incomplete	154	352	43.8	.00003

The frequency distribution of hit rates is shown in Figure 1. The theoretical distribution expected on the basis of chance is shown for comparison. The actual distribution was shifted to the right of the chance pattern.

Out of 288 completed tests, more tests had above-chance scores than below-chance scores: 138 had hit rates above chance, and 66 had hit rates below chance; 84 were at the chance level of 2/6. Thus, disregarding the tests that were at chance level, there were 204 tests that were above or below chance. Out of these 204 tests, according to the binomial distribution, the probability of tests having scores above chance was 0.476, or in other words 97 tests. In fact 138 were above chance, a highly significant difference ($P = 1 \times 10^{-8}$).

Overall Results: Tests With Two Callers

In tests with two callers, there were 745 trials altogether, with a hit rate of 55.2%, above the 50% expected by chance ($P = .003$). The effect size (Cohen's d) was 0.10.

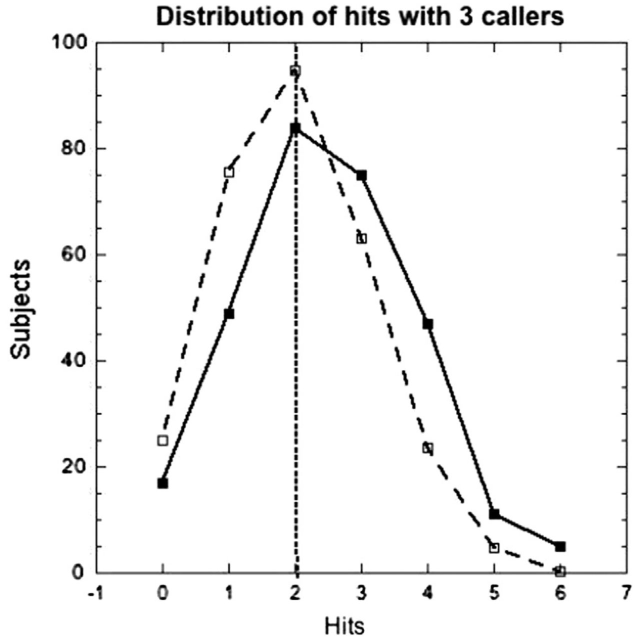

Figure 1. The number of subjects with hit rates from 0 to 6 in complete tests of six trials, each with three potential callers. The solid line shows the actual distribution of scores, and the dashed line the theoretical binomial distribution of results by chance. The vertical line indicates the chance score of 2/6.

Table 3. Telephone Telepathy Tests with Two Callers: Hit Rates in All Tests, Complete Tests, and Incomplete Tests

Tests	Hits	Trials	Hits (%)	*P*
All	411	745	55.2	.003
Complete	370	660	56.0	.001
Incomplete	41	85	48.2	NS

NS, nonsignificant.

There were 110 complete six-trial tests, with 660 trials, in which the hit rate was 56.0% ($P = .001$). The 95% confidence interval for this figure was from 52.3% to 56.8%. There were 42 incomplete tests, in which the hit rate was 41 out of 85 trials [48.2%; nonsignificant (NS)] (Table 3). In addition, 93 tests were registered but no trials were carried out, and hence there were no hits.

The frequency distribution of hit rates is shown in Figure 2. The theoretical distribution expected on the basis of chance is shown for comparison. As in Figure 1, the actual distribution was shifted to the right of the chance pattern.

Out of 110 completed tests, more tests had above-chance scores than below-chance scores: 49 had hit rates above chance, and 24 had hit rates below chance; 37 were at the chance level of 3/6. Thus, disregarding the tests that were at chance level, there were 73 tests that were above or below chance. Out of these 73 tests, the probability of tests having scores above chance was 0.5, or in other words 36.5 tests. In fact 49 were above chance ($P = .002$).

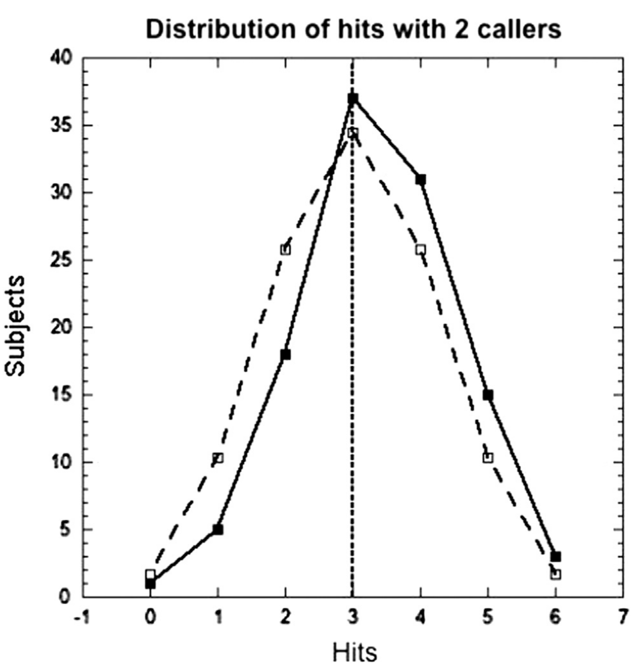

Figure 2. The number of subjects with hit rates from 0 to 6 in complete tests of six trials, each with two potential callers. The solid line shows the actual distribution of scores, and the dashed line the theoretical binomial distribution of results by chance. The vertical line indicates the chance score of 3/6.

Effects of Sex and Age of Subjects

In tests with three callers, and hence with a chance hit rate of 33.3%, combining the results from complete and incomplete tests, male subjects scored 452 hits out of 1063 trials (42.5%; $P < 1 \times 10^{-6}$) and female subjects 417 hits out of 1017 trials (41.0%; $P < 1 \times 10^{-6}$). The slightly higher hit rate with males was not significantly different from the hit rate with females.

In tests with two callers, with a chance hit rate of 50%, combining the results from complete and incomplete tests, male subjects scored 231 hits out of 414 trials (55.8%; $P = .01$) and female subjects 180 hits out of 332 trials (54.2%; $P = .06$). As in the tests with three callers, the slightly higher hit rate with males was not significantly different from the hit rate with females.

The hit rates for subjects in different age groups are shown in Figure 3, combining the results from complete and incomplete tests. In tests with three callers, the highest hit rates were with younger subjects in the 10−19 and 20−29 age groups, and the lowest hit rate was in the oldest age group. In the tests with two callers, the highest scores were in the 40−49 and 50+ age groups, although there were relatively few subjects in these groups. However, none of the differences among hit rates in different age groups was statistically significant.

Changes in Hit Rates Within Tests

In the complete tests, the hit rates on trials 1−6 were compared to see if there was any systematic pattern. In tests with three callers, there was no overall tendency for scores to increase or decrease throughout the tests (Figure 4). The lowest hit rate was on trial 4, but the difference between this hit rate and the hit rates in trials 3 and 5 was not statistically significant.

By contrast, in tests with two callers, there was a general tendency for hit rates to decline in the course of the test: the combined hit rate in the first three trials was 59.7% and in the second three was 53.7%, but this difference was not statistically significant. The lowest hit rate was on trial 2, and the difference between this hit rate and the hit rates in trials 1 and 3 was statistically significant ($P = .03$).

Effects of Repeated Testing of the Same Subjects

Some subjects were retested repeatedly, and took part in up to six successive tests. In Figure 5, we compare their average hit rates in these repeated tests. There was no obvious tendency for the scores to improve with practice, or to show a decline effect. In tests with three callers, the highest hit rates were on the first test. With two callers, hit rates on the first test were higher than in all the other tests except the third, where they were marginally higher than in the first test.

Effects of Time Delays Before Callers Responded

The randomly selected callers received a text message asking them to call the subject. Sometimes they responded almost immediately, but sometimes there were delays before they did so, and sometimes they failed to respond at all. When they did not respond the test itself stopped, and the test was incomplete. For completed tests, Table 4 shows the hit rates

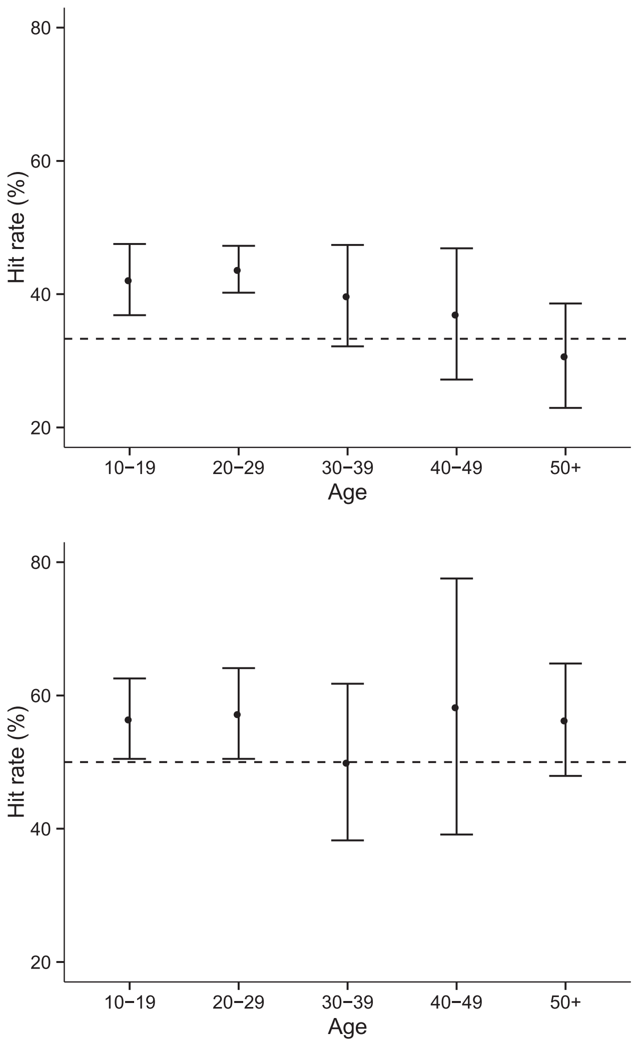

Figure 3. Hit rates of subjects in different age groups in tests with three (above) and two (below) callers, with 95% confidence intervals. The chance level of 33.3% in the three-caller tests is indicated by a horizontal dashed line, and of 50% in the two-caller tests by a horizontal solid line.

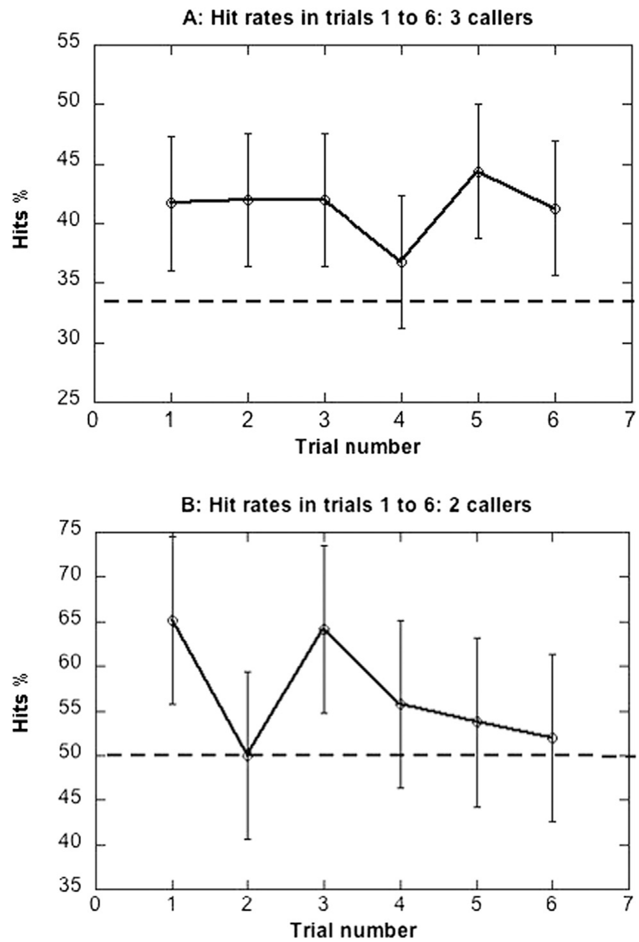

Figure 4. Hit rates in successive trials in complete tests with three callers (A, above) and with two callers (B, below), with 95% confidence intervals. Horizontal dashed lines indicate the hit rates expected by chance.

with different delays for tests with three callers. In general, the longer the delay, the higher the hit rate. But even in trials completed with delays of less than four minutes, the hit rate was very significantly above the chance level of 33.3%. The differences between hit rates with different delays were not significant statistically.

The effects of delays in callers' responses in tests with two callers are shown in Table 5. Again, the longer the responses were delayed, the higher the hit rate, but these differences were not significant statistically. However, even with no delays of more than four minutes, the hit rate of 55.3% was very significantly above the chance level of 50.0%.

Another way of looking at the effects of delayed responses is to compare complete tests in which all the responses of callers happened in less than four minutes, tests in which one or more of the trials had a caller delay of 4–10 min, and tests in which one or more trials had a caller delay of more than 10 min. The results are shown in Tables 6 and 7. Hit rates were higher in tests in which one or more of the trials had delays, and the longer the delays the bigger the effect. In tests in which all callers responded quickly, in less than four minutes, the hit rates were above chance, but *P* value was greater than 0.05. However, when some of the calls were delayed by more than four minutes but less than 10, the hit rates were above chance, with *P* values of 0.003 for tests with three callers and 0.03 for tests with two callers.

DISCUSSION

Hit Rates Compared with Previous Tests

These experiments showed that it is possible to do tests for telephone telepathy using an automated system involving mobile telephones under real-life conditions. The overall hit rates were positive and significantly above chance, as in previous research on telephone telepathy using landlines.[2–6]

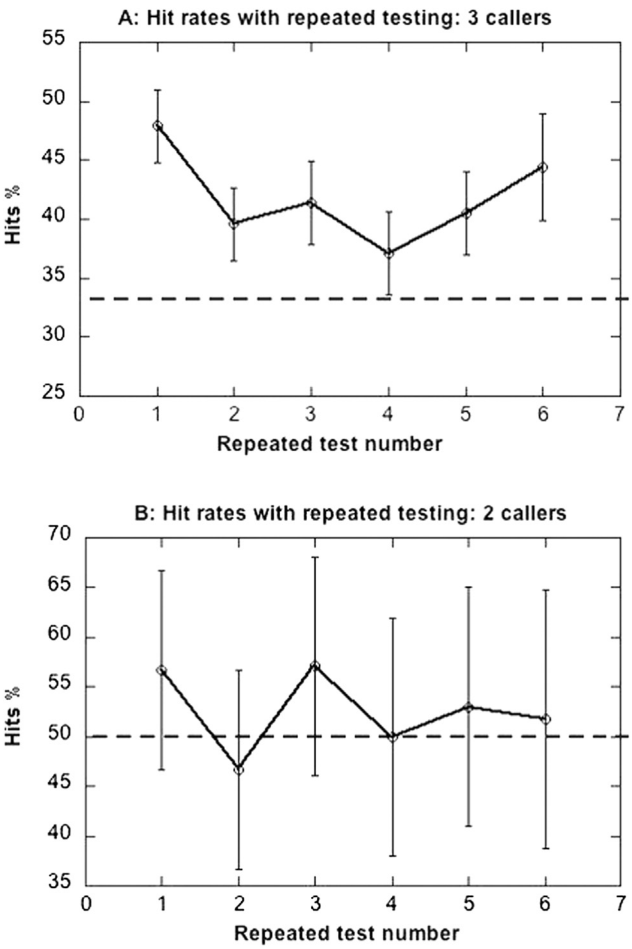

Figure 5. Effects of repeated testing on average hit rates in complete tests with three callers (A, above) and with two callers (B, below), with 95% confidence intervals. Horizontal dashed lines indicate the hit rates expected by chance. In the three-caller tests, the hits rates in tests 1 and 2 are out of a total of 240 trials; in test 3, 198; in tests 4 and 5, 186; in test 6, 126. In the two-caller tests, the hit rates in tests 1 and 2 are out of a total of 90 trials; in test 3, 84; in tests 4 and 5, 66; in test 6, 60.

The effect sizes (Cohen's *d*) for these automated tests with three callers were 0.19 and with two callers 0.10. These effect sizes were lower than in our earlier experiments on telephone telepathy, where they were 0.35 in unfilmed tests and 0.46 in filmed tests.[2,3] These differences may be explicable in terms of the way we recruited and interacted with subjects. For our landline experiments, we recruited subjects nationally through recruitment websites, and specified that we were looking for people who knew who was ringing before they picked up the phone, or who thought of themselves as intuitive. For our filmed experiments, we selected subjects who had high hit rates in preliminary unfilmed tests. In other words, we recruited potentially sensitive subjects, and for the filmed tests we selected subjects who had already had high hit rates in preliminary tests. Moreover, we interacted with the subjects personally, and also interacted personally with their callers, because we had to ring the randomly selected callers before each trial to ask them to call. This personal attention might have increased the callers' motivation to concentrate on the subject when they were making the call.

Table 4. Effects of Delays Before the Callers Responded on Hit Rates

Delay	Hits	Trials	Hits (%)	*P*
<4 min	466	1168	39.9	$<1 \times 10^{-6}$
4–10 min	134	304	44.1	$<1 \times 10^{-6}$
>10 min	115	256	44.9	$<1 \times 10^{-6}$

Data from completed tests with three callers.

Table 5. Effects of Delays Before the Callers Responded on Hit Rates

Delays	Hits	Trials	Hits (%)	*P*
<4 min	298	539	55.3	0.008
4–10 min	41	72	56.9	NS
>10 min	31	49	63.3	0.04

NS, nonsignificant.
Data from completed tests with two callers.

By contrast, in the tests described in this paper, we recruited subjects indirectly through research helpers, who were mostly students. These research helpers usually found it difficult to recruit subjects, because most people they asked responded by saying they were too busy. Although we asked the research helpers to recruit subjects who thought of themselves as intuitive, they often recruited anyone who was willing to take part. Then the subjects themselves had to recruit callers, who also often said they were too busy. Thus many of the subjects in the present experiment were probably not particularly sensitive, compared with subjects in our previous tests. Moreover, in these automated tests, where callers were responding to computer-generated requests, some may have been more perfunctory in the way they made the calls.

The Possibility of Cheating

In most previous studies on telephone, SMS, and email telepathy, subjects were filmed throughout the tests to ensure that they were not receiving any other information by tele-

Table 6. Tests in Which All Caller Responses Took Less than Four Minutes, Compared With Tests in Which at Least One Trial Involved a Delay of 4–10 min, and Tests in Which at Least One Trial Involved a Delay of More than 10 min

Delays in Trials Within Tests	Number of Tests	Hits	Trials	Hits (%)	*P*
All <4 min	56	117	336	34.8	NS
Some 4–10 min	77	188	462	40.7	.0003
Some >10 min	155	410	930	44.1	$<1 \times 10^{-6}$

NS, nonsignificant.
Data from completed tests with three callers.

Table 7. Tests in Which All Caller Responses Took Less than Four Minutes, Compared with Tests in Which at Least 1 Trial Involved a Delay of 4–10 min, and Tests in Which at Least 1 Trial Involved a Delay of More than 10 min

Delays in Trials Within Tests	Number of Tests	Hits	Trials	Hits (%)	*P*
All <4 min	47	150	282	53.2	NS
Some 4–10 min	37	126	222	56.8	.03
Some >10 min	26	94	156	60.3	.006

NS, nonsignificant.
Data from completed tests with two callers.

phone, text message, or email.[3–7] In the present study, most of the tests were unsupervised, leaving open the possibility that some of the subjects could have cheated. However, we think cheating was very unlikely for two reasons. First, there was no incentive to do so. Research helpers were paid for the number of people they recruited, not according to hit rates. And second, if some subjects were cheating, their scores would have distorted the pattern of scores; presumably cheaters would have had high scores, and most people would have scored at the levels expected by chance. Yet in fact many people's scores were above chance. In tests with three callers, 138 were above the chance level, compared with 61 below chance. In tests with two callers, 49 were above chance and 24 below. Moreover, Figures 1 and 2 show that the overall pattern of scores is shifted to the right compared with the chance distribution in a way that does not suggest a minority of subjects were cheating. If cheating was occurring, instead of a few subjects cheating a lot, most subjects would have had to cheat a little, and to do so they would have had to be in collusion with at least one of their callers. This seems very implausible.

Improvements or Declines?

Do subjects improve with practice? When we looked at hit rates in the six successive trials within a test, there was no clear pattern (Figure 4). In tests with two callers, the highest hit rate was on the first trial, with a tendency to decline thereafter, and in tests with three callers, the highest hit rate was on the fifth trial, showing a slight tendency towards improvement. When the same subjects were tested repeatedly, in all repeat tests with three callers and in all except the third test with two callers, the hit rates were lower than on the first test (Figure 5). Thus under the circumstances of our tests, practice did not lead to improvements in hit rates.

We started this investigation using tests with three callers, and then changed to doing tests with two callers, largely because it was easier to organize tests involving fewer people. The tests with three callers showed a stronger positive effect: a hit rate 41.8%, as opposed to 33.3% expected by chance ($P < 1 \times 10^{-15}$). In tests with two callers, the hit rate was 55.2% as opposed to 50% expected by chance ($P = .003$). The effect sizes (Cohen's *d*) were 0.19 and 0.10, respectively. Was this because the subjects found it easier to guess correctly when there were more callers to choose from, or was it because the callers were more effective at communicating their intention when they called less often?

In many fields of scientific research, effects seem to decline with time. This puzzling phenomenon was first discovered in parapsychology in the 1930s and has since been documented in other fields, including psychology, biology, and medicine.[12] We looked at our own data to find out if there had been a decline in hit rates in the course of our own series of tests, but found that there was no clear overall pattern of decline, just as there was no clear pattern of decline (or increase) when we looked at the data for individual subjects who were tested repeatedly both within tests (Figure 4) and in repeated tests (Figure 5), as discussed above.

Delayed Responses by Callers

In both series of tests, there was a tendency for hit rates to increase when callers took longer to respond to the request to call the subject (Tables 4 and 5). This could suggest that delays gave subjects clues about who the callers might be. For example, they might know that one of their callers was often very busy or was generally slow to respond to text messages, so a long delay might cause them to guess that this person was the caller.

However, this hypothesis cannot explain all the effects that we observed. The subjects were told that when one trial had ended, the next would begin after a random time delay, but they were not told how long this delay might be. Also, they were not aware when the first trial would happen, because this too depended on a random time delay. They would be able to estimate possible lengths of delays between trials only after they had done several trials, and even then this estimate would be only approximate since the delays were randomized. And since the randomized delays could be as long as 10 min before the next trial began, if they received a call more than four minutes but less than 10 min after the previous trial, this could not give reliable information about the identity of the caller. For example, a caller might have responded immediately, with no delay, to instructions to call the subject nine minutes after the previous trial, or a caller might have taken eight minutes to respond to a request made one minute after the previous trial. From the subject's point of view, both calls would come 9–10 min after the previous trial, and hence the quick and slow responders would not be distinguishable.

It is possible that when delays in callers' responses were greater than 10 min, some subjects might have obtained clues that influenced their guessing, and hence inflated their hit rates in a way that was not telepathic. This delay hypothesis might explain some of the above-chance hit rates, but it cannot explain them all, for two main reasons. First, subjects did not know when the first trial in the test would take place, and thus had no way of knowing whether the time lag before the test started was due to the caller or to the system. Yet hit rates on the first trials in these tests were well above chance (Figure 4). Second, if subjects learned through experience to expect time delays between trials of less than about 10 min, then their hit rates should have increased as tests went on, and they should also have performed better and better when they were tested repeatedly. Yet the data do not show a progressive improvement (Figure 5).

There could also be a completely different explanation for the enhancement of hit rates by delays in callers' responses. Maybe the subjects' ability to respond telepathically was reduced by testing them too frequently. A suggestion that this might be the case came from automated online telepathy tests in which speeding up the procedure led to reduced hit rates.[13] When tested too frequently, subjects may become more self-conscious about their feelings and thoughts, and this self-consciousness might inhibit their telepathic sensitivity. Instead of trusting their feelings or intuitions, they may become doubtful about their abilities, and afraid of guessing wrongly. Such thoughts and worries do not enhance telepathic sensitivity, but interfere with it. With longer delays between trials, they have longer to get absorbed in other activities and stop thinking about the test in which they are participating. If this is true, then increasing the time between trials by setting the time delay to, say, 20–30 min should result in increased hit rates. This is a testable hypothesis.

Possible Improvements in Future Studies

Our investigations show, for the first time, that it is possible to carry out automated telephone telepathy tests using mobile telephones. The main problems our study has highlighted that need to be addressed in future projects are as follows:

1. The difficulty of recruiting participants. We have so far found only three ways of motivating people to take part in these and other telepathy tests:

i) paying the participants,
ii) paying people to recruit participants, as we did in this experiment, and
iii) encouraging students to run an experiment as a project, in which case their friends and family were motivated to help them.

There could be more effective ways of motivating people to take part in tests. One method might be to make the tests part of a training program for enhanced psi abilities. Simply doing these tests repeatedly does not show a clear pattern of learning and improvement (Figure 5), so repeated participation in tests alone is inadequate. The tests would need to be embedded in an intuition training program, and they would give a quantitative measurement of how effective such a program was. If such a program could be combined with telephone telepathy tests in the form of an app or an online course, then, if it worked, it might be possible to encourage large-scale participation.

Perhaps the most effective way to create incentives for people to do telepathy tests, and to take part in intuition training programs, is to reward good telepathic performance. For example, if it were possible to train people to raise their scores in automated telephone telepathy tests, then it might be possible to run telepathy competitions under cheat-proof conditions, with prizes. As in competitive sports events, like gymnastics, competitions would provide strong incentives for people to develop their skills. Eventually there could even be regional or national telepathy contests, with the finals shown on television. Ultimately, these contests could lead to the establishment of telepathy Olympics. If this were to happen, public and academic attitudes to psi abilities would undergo a dramatic shift.

2. In precognition tests, it is relatively easy to design cheat-proof procedures, because people's responses can be recorded before the target is determined at random. If the responses are recorded automatically along with the time at which they are made, the evidence for any precognitive effects could be very strong. By contrast, telepathy tests inevitably involve other people, who need to know when to contact the subjects. Meanwhile the subjects must not know by normal means who is about to contact them. This leaves open the possibility of a caller contacting a subject by another means, for example through a text message or email, enabling the subject to cheat. We think that cheating was extremely unlikely under our test conditions, and the pattern of results is not consistent with cheating, as discussed above. But unless the subjects are filmed or otherwise supervised, the possibility of cheating cannot be ruled out completely. And when the subjects are filmed, the difficulty of recruitment is made even greater. Also, the advantage of a test in real-life conditions is lost as far as the subjects are concerned, because they have to stay in a particular place where they can be securely monitored. It should be possible to devise new versions of automated telephone telepathy tests that greatly reduce the possibility of cheating, but it would be hard to eliminate it altogether, and hence to eliminate the requirement for filmed tests under secure conditions.

3. A problem with the procedure we adopted in the present experiments was the variable time taken for callers to respond to the text message asking them to call the subjects. Many responded quite promptly, but on some occasions, some callers responded with delays of 10 min or more, which could have given the subjects clues as to who the callers were. The fact that trials in which there were long delays before the caller responded showed higher hit rates supports this possibility (Tables 4–7). In designing future tests, it is important to find methods that do not allow for caller-dependent delays. One possible procedure is to have all participants linked up through a telephone system continuously during test periods, as in a conference call system, so that callers are able to respond rapidly when requested to do so.

4. A big problem in testing psi abilities is that the detection of psi may take place unconsciously or physiologically, whereas in order for subjects to give verbal or numerical responses they have to engage the conscious mind. Perhaps in automated tests this problem could be reduced by displaying pictures of the callers, and asking people to indicate who is calling by touching a picture on a screen.

As mobile phones and other portable electronic devices become ever more widely used, the opportunities for psychological research in general, and psi research in particular, will increase greatly. We hope that our experiences with these automated telephone telepathy tests will contribute to these developments.

Acknowledgments

We thank Matous Novak, of Mobifi Ltd., for his work in programming the test system and database, the research helpers who recruited participants for these tests, the participants themselves, and Jan van Bolhuis for his advice on statistical analysis. We gratefully acknowledge financial support from the Perrott–Warrick Fund at Trinity College, Cambridge, Mr Addison Fischer, the Planet Heritage Foundation, the Institute of Noetic Sciences, and the Watson Family Foundation.

REFERENCES

1. Sheldrake R. *The Sense of Being Stared At, And Other Aspects of the Extended Mind.* New York, NY: Crown; 2003.
2. Sheldrake R, Smart P. Experimental tests for telephone telepathy. *J Soc Psychical Res*. 2003;67(3):184–199.
3. Sheldrake R, Smart P. Videotaped experiments on telephone telepathy. *J Parapsychology*. 2003;67(Spring):147–166.
4. Sheldrake R, Godwin H, Rockell S. A filmed experiment on telephone telepathy with the Nolan sisters. *J Soc Psychical Res*. 2004;68(3):168–172.
5. Lobach E, Bierman DJ. Who's calling at this hour? Local sidereal times and telephone telepathy. *Proceedings of Parapsychology Association Annual Convention, Vienna;* 2004:91–97.
6. Schmidt S, Erath D, Ivanova V, Walach H. Do you know who is calling? Experiments on anomalous cognition in phone call receivers. *Open Psychol J*. 2009;2:12–18.
7. Sheldrake R, Smart P. Testing for telepathy in connection with e-mails. *Percept Mot Skills*. 2005;101(3):771–786.
8. Sheldrake R, Avraamides L. An automated test for telepathy in connection with emails. *J Sci Exploration*. 2009;23(1):29–36.
9. Sheldrake R, Avraamides L, Novak M. Sensing the sending of SMS messages: an automated test. *Explore*. 2009;5(5):272–276.
10. Haraldsson E. Representative national surveys of psychic phenomena: Iceland, Great Britain, Sweden, USA and Gallup's multinational survey. *J Soc Psychical Res*. 1985;53(801):145–157.
11. Haraldsson E, Hootkooper JM. Psychic experiences in the multinational human values study: who reports them. *J Am Soc Psychical Res*. 1991;85(April):145–165.
12. Schooler J. Unpublished results hide the decline effect. *Nature.* 2011;470(7335):437.
13. Sheldrake R, Beeharee A. A rapid online telepathy test. *Psychol Rep*. 2009;104(3):957–990.