

CORRESPONDENCE

To the Editor,

In a recent publication entitled "A Filmed Experiment on Telephone Telepathy with the Nolan Sisters", Rupert Sheldrake and colleagues report an experiment conducted with the Nolan Sisters in 2003 during a television show, designed to test the hypothesis of telepathy by telephone (Sheldrake, Godwin & Rockell, 2004). According to their statistical analysis of the results of a few trials, the authors conclude "*that the results support the hypothesis of telepathic communication*" (p.172). I should like to discuss here why this conclusion *is*, in my view, mistaken.

Let us briefly recall that in the experimental procedure described in the original publication, Colleen, one of the five Nolan sisters, had to guess, just before picking up the phone, who among her four sisters was calling her. The experiment was filmed and, according to the authors, any "*'normal' sensory clues*" were prevented. 12 successive trials were planned, separated by five-minute intervals, the calling sister for each trial being randomly selected by the throw of a die.

Colleen correctly guessed who was calling in 6 out of the 12 trials, leading the authors to the above conclusion. This above-chance result was claimed to be significant at the $p = 0.05$ level, and consequently the null hypothesis (i.e. that the results are obtained by chance) was rejected in favour of the hypothesis of the highly "extraordinary" and controversial telepathy phenomenon. However, using the exact binomial test (which has to be preferred here over other statistical tests since within the null hypothesis the results follow exactly the well-known binomial law), the correct value corresponding to a 6/12 score is in fact $p = 0.0544$ (or 1/18). Since this value is strictly greater than 0.05, it is incorrect to conclude that "the results [were] significant at the $p = 0.05$ level" and it is misleading to talk about "Colleen Nolan's above-chance success rate" (p.171).

Moreover, there are some mistakes when the authors write "The positive result could simply have arisen by chance, but the odds against this explanation are 19:1" (p.171). This value is incorrect, since the true p value must be used, leading in fact to the odds against the explanation by chance of 17:1. Rejecting the null hypothesis because the odds against it are 19:1 is already "risky" (though it seems to be the standard in parapsychology and other social sciences), but with the odds 17:1 it is even more venturous! The second point, which is more conceptual, is that the p value of a statistical test is *not* the probability for the null hypothesis (i.e. chance) to be true. The p value gives the probability to obtain, within the null hypothesis (by chance), a score greater or equal to the one observed (for a one-tailed test, which is the case here). This is the only interpretation that can be made, even if of course, the smaller the p value is, the more reasonable it is to think that the null hypothesis is not the correct explanation.

Furthermore, Sheldrake et al. write in their article that "on two occasions Colleen picked up the phone before making her guess, contrary to her

instructions" (pp.170–171). This is clearly a violation of the experimental procedure and the authors agree that "these two trials should be excluded from the total". Nevertheless, they do not reconsider the statistical analysis, arguing that with 10 or 12 trials the success rate remains the same: 5 out of 10 instead of 6 out of 12. Using this spurious argument, the authors stick to their conclusion in their abstract that "the results are significant at the $p = 0.05$ level" and do not reject expressions like "positive result" or "above-chance success rate" as they should have done. It is indeed obvious that the same success rate can be significant for a given number of trials while it is non-significant for another (smaller) number of trials! In this case, although a score of 6/12 was almost significant (p just over 0.05) a score of 5/10, corresponding to the same success rate, is far from being significant at the 0.05 level. Calculations using the exact binomial test show that for such a result, the p value is $p = 0.0781$. This value would have given the odds against the chance explanation (setting aside the fact that this interpretation of the p value is erroneous, as discussed above) of 12:1.

This shows that, with a more careful statistical analysis, Sheldrake et al. should have come to this kind of conclusion, stated in a more rigorous way: "the odds against the fact that chance would give such a high score for this experiment are 12:1". The reader could then have made up his or her mind concerning the soundness of the conclusion that "the results support the hypothesis of telepathic communication".

Lastly, the authors themselves admit that cheating was possible (for instance using electronic devices). Although the authors are "confident" on this point and claim that "it is not plausible", cheating cannot be totally ruled out. For the sake of scientific rigour, an experiment where any form of "normal" communication is really made impossible would have been preferable; otherwise any sound conclusion can unfortunately not be drawn from the results.

LPMCEN

FLORENT TOURNUS

Université Claude Bernard Lyon 1
 Domaine scientifique de la Doua
 Bâtiment Léon Brillouin
 43 Bd du 11 Novembre 1918
 F-69622 Villeurbanne, FRANCE

REFERENCE

- Sheldrake, R., Godwin, H. and Rockell, S. (2004) A filmed experiment on telephone telepathy with the Nolan Sisters. *JSPR* 68, 168–172.

Rupert Sheldrake replies:

Dr Florent Tournus is correct in pointing out the limitations of a study with a small number of trials. However, he ignores the context of this test, which was a small-scale replication, carried out for television, of an experimental investigation that had already been performed on a much larger scale, with 271 trials (Sheldrake & Smart, 2003). In these previous videotaped tests, there were 122 hits (45%), compared with a 25% hit rate expected by chance

April 2007]

($p = 1 \times 10^{-12}$). In the replication under discussion here, there were only 12 trials. The hit rate was 50%, slightly higher than in the previous test, but because of the small sample size, the results were inevitably much less significant statistically.

Tournus is right in saying that by the binomial test a hit rate of 6 out of 12 gives $p = 0.054$. By normal convention, practised in all branches of science, values between 0.046 and 0.054 are rounded to 0.05.

As we pointed out in our paper, in two trials Colleen picked up the phone before making her guess. This was contrary to her instructions, because it is conceivable that a message could have been passed to her from the caller. When excluded, the hit rate is 5 out of 10 (still 50%), but of course the significance drops, as Tournus points out. However, both Colleen and the caller were continuously observed directly and also filmed, and there was no sign of any message being passed between them, either by direct observation at the time or by subsequent examination of the videotape, so these trials are probably valid. In any case, one of them was a miss.

It is conceivable, as we pointed out, that the participants could have been cheating, for example if they had electronic devices concealed in their clothing. But even if we had strip-searched them to rule out such a possibility, extreme sceptics would still be able to argue that they could have had small radio devices implanted in their teeth, or subcutaneously. But we think cheating is highly unlikely because the Nolan sisters would have no motive to do it. They did not know about the design of the experiment in advance, nor was the choice of which sister would serve as the subject determined beforehand. If they had been skilled stage magicians with a very strong desire to cheat, perhaps they could have done so with the aid of miniaturised electronic devices and great skill, but this seems to us implausible.

In the context of previous large-scale tests with highly significant positive results, the results of this small-scale replication support the hypothesis of telepathic communication. We do not claim that they *prove* it. Tournus implies that we were somehow inhibiting readers from making up their own minds by providing our own interpretation. It was our intention that any reader who wanted to form a considered judgement should be able to look at the detailed results of this test in the context of the larger experiments of which it was a replication. This small study was published as a *Note* rather than a full-scale paper in the *JSPR* precisely because of its limited number of trials. We concluded, "of course, further replications will be needed". This is hardly the exaggerated claim that Tournus implies we were making.

20 Willow Road
London NW3 1TJ

REFERENCE

- Sheldrake, R. and Smart, P. (2003) Videotaped experiments on telephone telepathy. *JP* 67, 147-166.