THE SENSE OF BEING STARED AT:
DO HIT RATES IMPROVE AS TESTS GO ON?

by Rupert Sheldrake

ABSTRACT

Simple experiments on the sense of being stared at have given repeatable, positive results that are highly significant statistically. In these experiments, people work in pairs. The staree sits with his or her back to the starer, who either looks at the back of the staree's neck, or looks away, in a random sequence. In each trial, the staree has to guess whether or not the starer is looking. However, when Marks and Colwell (2001) and Lobach and Bierman (2004) conducted tests of this kind, some of their experiments gave results not significantly different from chance, and they attempted to explain the positive results in staring tests as artifacts. Their hypotheses predict that positive scores should arise only in trials with feedback, only in trials with one particular kind of randomization, and that scores should increase towards the end of the experimental session. I have examined the data from the first and second halves of more than 19,000 trials to test these predictions. Both with and without feedback, and also with different randomization methods, the scores were positive and statistically significant in both the first and the second halves of tests. With feedback there was a small increase in scores in the second halves, but this was not statistically significant. Without feedback, there was a tendency for the scores to decline. In a trial-by-trial analysis of one large-scale experiment, the highest hit rate occurred in the very first trial for starees who were about to receive feedback, before any feedback had actually been given! Thus the beneficial effect of feedback may depend not so much on the feedback itself as on the state of mind of the participants.

INTRODUCTION

Most people have had the experience of turning round with the feeling that someone is looking at them from behind, finding that this is the case. Most people have also found that they can sometimes make people turn around just by looking at them (Sheldrake, 1994, 2003a; Cottrell, Winer & Smith, 1996).

The sense of being stared at (or *scopaesthesia* — Carpenter, 2005) can be investigated experimentally with people working in pairs, with the starer sitting behind the staree. In a randomized series of trials, the starer either looks at the back of the staree's neck, or looks away and thinks of something else. The starer signals to the staree when the trial begins by means of a mechanically produced sound. After about 10 seconds the staree guesses out loud whether or not he or she is being looked at, and the starer records whether the guess is right or wrong (Sheldrake, 1999). By chance, 50% of the guesses would be correct.

Experiments of this kind are so simple that students can do them in classroom experiments or as projects. Many thousands of trials of this kind have been carried out in schools and colleges, with highly significant positive results. I have organized many of these trials myself, both with children and with adults, and they have been replicated independently by dozens of other investigators in several different countries (Sheldrake, 1998, 1999, 2003a, 2005).

The results usually showed a distinctive pattern: in the 'looking' trials,

scores were above the chance level, usually around 60%, while in the 'not looking' trials they were close to the chance level of 50%, with overall success rates around 55% (Sheldrake, 1999, 2003a). The pattern of results was similar whether or not the starees were blindfolded or received trial-by-trial feedback (Sheldrake, 2001), and when they were looked at through closed windows (Sheldrake, 2000) or through a one-way mirror (Colwell, Schröder & Sladen, 2001: Experiment 1).

However, not all studies have given similar results. After obtaining highly significant positive results in their first experiment for trials in which feedback was provided, Colwell et al. (2001) conducted a second experiment in which they changed two conditions: a different person served as starer and they used a different randomization procedure. The results were close to the chance level. They concluded that the difference from their first experiment must be due to the randomization method rather than the change of starer. But their assumption that the difference in outcome could not be due to a change in starer may well be invalid: in a study using a different technique, in which starees were looked at through a closed circuit television (CCTV), Wiseman and Schlitz (1997) found a striking experimenter effect whereby the results were non-significant when Wiseman, a sceptic, was the starer, while with Schlitz as starer, the results were positive and statistically significant.

Marks and Colwell (2001) argued that the positive result in Colwell et al.'s (2001) first experiment might be an artifact that arose because of a particular set of randomized score-sheets that I used in some of my experiments, which Colwell et al. also used in theirs. The randomizations involved counterbalanced sequences, and Marks and Colwell showed that these particular sequences contained more alternations than would be expected in 'structureless' randomizations. They argued that when starees were given trial-by-trial feedback, they could have learned implicitly that there were patterns in the randomizations, and somehow used this information to score at above-chance levels. This hypothesis implies that scores should increase towards the end of tests because there has been more opportunity for implicit learning to occur.

Lobach and Bierman (2004) carried out two staring experiments with student participants in which starers and starees were separated by a one-way mirror (they also carried out a study using CCTV). In the first experiment, there was a significant positive result only with starers who were sceptics, and not with starers who were believers. In the second experiment there was no significant effect. They discussed several reasons for the weakness or absence of effects in their studies, including a possible inhibitory effect of the formal laboratory setting and the fact that starees were not blindfolded but had their eyes open and had to respond by making their guesses through a computer. Like Marks and Colwell (2001), they also explored the possibility that, in some of my own tests, starees given trial-by–trial feedback could have obtained hit rates above chance artifactually, through a combination of two effects. First, with randomization sequences that had too frequent alternations between looking and not-looking trials, starees could have elevated their hit rates by alternating their guesses. Second, if the total numbers of looking and not-looking trials were equal, or nearly equal, and if subjects remembered the number of looking and not-looking trials, then towards the end of the test it should become increasingly

possible to predict whether a looking or not-looking trial would come next. Scores should therefore improve towards the end of the test. They carried out computer simulations that showed that such guessing strategies could indeed produce artifactual positive results. Moreover, by adding in a response bias in favour of saying "looking", their simulations reproduced a similar pattern of results to those found in many previous staring tests, with hit rates in 'looking' trials close to 60% and in 'not-looking' trials about 50%.

Thus both the Marks and Colwell hypothesis and the Lobach and Bierman hypothesis would predict that scores should improve towards the end of tests in which feedback was given.

By contrast, in a small-scale study with trial-by-trial feedback, Radin (2004) found the opposite: scores were highest in the first few trials and fell off towards the end of the session.

For all these reasons, it is of interest to know whether there is a general tendency for scores to increase or decline during tests on the sense of being stared at. I therefore re-examined the data from more than 19,000 trials conducted by myself and by other investigators to see if scores increased or decreased as the sessions went on. I did this by comparing the results from the first and second halves of the sessions, which usually consisted of 20 trials altogether. I analyzed results from tests conducted with and without feedback, and from tests using different randomization procedures: counterbalanced randomization sequences, "structureless" randomizations given by coin-tossing and computerized randomizations.

These comparisons enabled several different hypotheses to be tested:–

1. If the positive results in staring trials depend on the Lobach and Bierman (2004) strategy or the implicit learning hypothesis of Marks and Colwell (2001), results should increase significantly in the second half of the tests. This improvement should only be apparent in experiments with trial-by-trial feedback, and with counterbalanced randomization sequences. Both these hypotheses would predict that in experiments with no feedback and with 'structureless' randomization, results should be at chance levels throughout.

2. If the positive results in these trials are a result of artifacts arising from subtle sensory cues that are learned as a result of feedback, then scores should be higher in the second half of trials with feedback than in the first half, and should remain at chance levels in trials without feedback. A similar pattern of results should be obtained whatever the randomization system.

3. If the decline effect observed by Radin (2004) is of general occurrence, then scores should be lower in the second halves of the sessions.

In order to test these predictions, I have compared the results from the first and second halves of the test sessions.

METHODS

I used data from my own tests and from those conducted by other investigators, details of which have already been described in previous publications. I retrieved the original score sheets for all these tests, and examined these score sheets to find out how many guesses were correct in the first and second halves of the test session for each subject. These 'first' and 'second' scores were then tabulated.

Both the Marks and Colwell and Lobach and Bierman hypotheses applied specifically to a particular set of counterbalanced randomized sequences that I used in some of my own experiments, and which were also used in tests conducted in schools in Connecticut by schoolteachers. Also, their hypotheses apply only to tests in which trial-by-trial feedback was given. For this analysis I selected experiments that included these particular sequences, and compared them with other experiments in which different randomization methods were used and/or there was no feedback. In particular, I compared experiments in Connecticut schools with the counterbalanced randomized sheets and with coin tossing as the means of randomization. I compared experiments in a London school using counterbalanced randomization sheets with and without feedback. I also included an experiment I conducted with adult participants with coin-tossing randomization and feedback.

I also included data from an experiment conducted with automated instructions online in which the randomization was provided by an automatic randomization system, as described by Sheldrake et al. (2008). As in the coin-tossing method, the number of looking and not-looking trials in each test varied at random, although on average there were roughly equal numbers of each. In these tests with automated instructions, some starees received feedback and others did not. As described by Sheldrake et al. (2008), data from high-scoring participants (those with more than 17 hits out of 20) were not included in the analysis because of the possibility that they might have been cheating.

In this paper I describe the results of all these pre-selected data-sets—there was no exclusion of data-sets after the analysis had been conducted. The details were as follows:−

1. *Connecticut 1996 (CT-96):* Experiments with feedback in five schools in Connecticut, USA, carried out by science teachers using coin-tossing randomization (Sheldrake, 1998, Table 2). In these tests the number of trials per session differed from school to school, and ranged from 20 to 40. There were always even numbers of trials. For the analysis reported here, the set of trials was divided into two equal halves. When there were 40 trials per session, for example, the first sample consisted of the first 20 and the second of the remaining 20 trials.

2. *Connecticut 1997 (CT-97):* Experiments without feedback in seven schools in Connecticut using my score sheets with counterbalanced randomizations and 20 trials per session. In these and other tests using my counterbalanced randomizations there was a set of 24 different score sheets. These tests were carried out by science teachers in seven schools (Sheldrake, 1999, Table 3, excluding data from Sam Brown).

3. *University College School, London, Winter 1997 (UCS-F):* My own experiments with children given feedback, using score sheets with counterbalanced randomizations and 20 trials per session (Sheldrake, 2001, Table 1A).

4. *UCS Junior Branch, Spring 1997 (UCS-N):* My own experiments with children without feedback, with starers and starees separated by closed windows, using score sheets with counterbalanced randomizations and 20 trials per session (Sheldrake, 2000, Table 1).

101

5. *Holma College, Höör, Sweden, 2000 (Holma):* My own experiments with adults given feedback, using coin-tossing for randomization and 20 trials per session (Sheldrake, 2002).

6. Data from the automated online test procedure accessed through my web site (www.sheldrake.org) as described in Sheldrake et al. (2008).

The statistical significance of the different hit rates was calculated using the binomial test, one-sided, with the null hypothesis that the hit rates would be at the chance level of 0.5. For a comparison of the hit rates in the first and second halves of each experiment, the 2×2 chi-squared test was used (Campbell, 1989).

RESULTS

Tests With Feedback

In all the tests with feedback, the hit rates were above chance in both the first and the second halves. In the three conventional experiments, there was a slight increase in the hit rates in the second halves, but there was no such increase in the automated test online (Table 1A). None of these changes in hit rate was statistically significant.

Table 1

Comparison of hit rates in the first and second halves of test sessions with feedback (A) and without feedback (B). The total number of trials in each half is shown, together with the numbers and percentages of hits in the first (1) and second (2) halves. There were three methods of randomization: by tossing coins (Coins), using counterbalanced randomization sheets (Sheets) and by an automatic computer-based randomization system (Auto). The final column contains p values from statistical tests for a difference between 1 and 2.

A: With Feedback

Test	Random-ization	Trials	1. Hits	p	2. Hits	p	1. %	2. %	p diff
CT-96	Coins	822	444	0.01	464	3×10^{-4}	54.0	56.4	0.32
UCS-F	Sheets	780	446	3×10^{-4}	461	$<1 \times 10^{-6}$	57.2	59.1	0.44
Holma	Coins	380	206	0.06	219	0.002	54.2	57.6	0.34
Online	Auto	2880	1622	$<1 \times 10^{-6}$	1617	$<1 \times 10^{-6}$	56.3	56.2	0.89

B: Without Feedback

Test	Random-ization	Trials	1. Hits	p	2. Hits	p	1. %	2. %	p diff
CT-97	Sheets	940	522	4×10^{-3}	517	0.001	55.5	55.0	0.81
UCS-N	Sheets	1540	820	0.005	796	0.09	53.2	51.6	0.22
Online	Auto	2530	1326	0.008	1319	0.02	52.4	52.1	0.84

Tests Without Feedback

In the experiments without feedback, the scores were above chance levels in both the first and the second halves (Table 1B). In all cases, the percentage

of hits was slightly lower in the second halves of the tests, but the differences were not statistically significant.

Trialwise Results

For the automated tests, the trial-by-trial results for tests with and without feedback are shown in Figure 1. The hit rates with feedback were generally higher than those without, but there was no obvious trend over time, nor was any significant trend revealed by linear regressions. The most striking feature was the data for the first trial, where starees who were about to receive feedback had an extraordinarily high hit rate of 67.7%, while those with no feedback were slightly below the chance level, with 47.8%. This difference was very significant statistically (chi-squared = 21.9; $p \ll 1 \times 10^{-6}$).
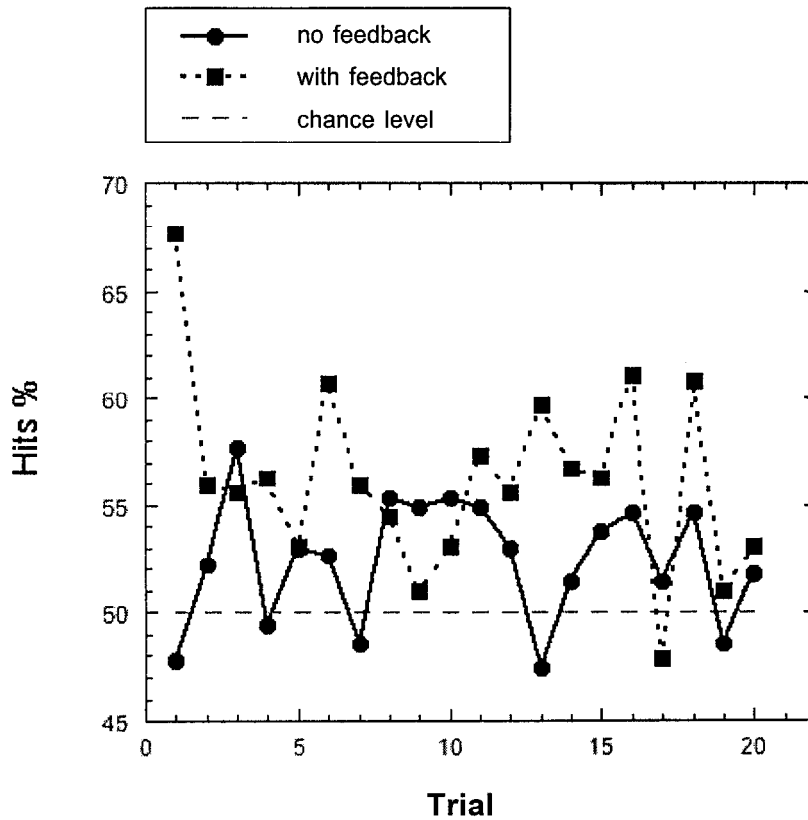

Figure 1. Trial-by-trial hit rates in the experiment with automated instructions online for tests with and without feedback.

DISCUSSION

The overall results were clear. In all experiments, the scores were very significantly above chance in both the first and the second halves of the tests.

In trials with feedback, there was a small increase in scores in the second halves of the sessions, but this was not statistically significant. In trials without

feedback, there was a small but non-significant decrease in scores in the second halves (Table 1). These general conclusions are illustrated in more detail by the trialwise data from the test with automated instructions online (Figure 1).

These findings have the following implications for Hypotheses 1 to 3 outlined in the Introduction:—

1. The hypothesis of Marks and Colwell (2001) predicted that scores would be above chance in the second halves, but only in trials with feedback and with counterbalanced randomizations. This is clearly not the case. The scores in the first halves were also above chance, and there was no significant increase in the second halves, even with feedback (Tables 1A & B). Moreover, the results were very similar with counterbalanced randomizations and with 'structure-less' randomizations produced by coin tossing. Thus the data clearly go against the Marks and Colwell hypothesis, and also the Lobach and Bierman (2004) hypothesis, in so far as it predicts that scores should increase towards the end of the test. However, the Lobach and Bierman hypothesis would allow starees to obtain above-chance scores even in the first half by using an alternating guess strategy, but this would only be possible with the counterbalanced sequences and with feedback. The experiment that met these conditions was UCS-F (Table 1A), and if these data were taken in isolation, their hypothesis might seem plausible. But it is not plausible, in view of the fact that similar patterns of results were obtained with other randomization methods and in tests without feedback.

2. The hypothesis that subjects learned through feedback to recognize subtle sensory clues likewise conflicts with the fact that the scores were positive both with and without feedback. Moreover, the scores were positive in the first halves of the sessions, before this hypothetical learning could plausibly have happened. Most remarkably of all, the highest hit rate occurred in the very first trial for starees in the feedback condition (Figure 1). But precisely because this was the first trial, when they made their guess they had not received any feedback at all! The large and very significant contrast between subjects in the feedback and no-feedback conditions cannot be explained in terms of feedback or its absence. Perhaps the most plausible explanation would be in terms of psychological differences caused by these different situations: starees knowing they are about to receive feedback may be more engaged with the testing process, and less nervous. By contrast, the lack of feedback may have an alien-ating effect on the starees and make them more self-conscious, inhibiting their ability to guess correctly. But whatever the explanation, it is clear that the higher hit rates with feedback than without are not because of any learning effect enabled by the feedback, because feedback confers an advantage right from the outset, even before it has happened.

3. The decline effect observed by Radin (2004) in tests with trial-by-trial feedback was not observed here and hence does not seem to be a general phenomenon in trials of this kind. Radin's subjects were tested under formal laboratory conditions, and about half of them were teenagers unused to this kind of environment. Radin himself has suggested that they might have been intimidated by being tested in a laboratory known for investigating psychic phenomena, and may have experienced rising performance anxiety (Radin, personal communication, 6 June 2004). The subjects in the experiments

described in this paper were tested under more informal conditions, where their enthusiasm and curiosity probably sustained the interest of both starees and starers.

If the sense of being stared at is real, then it should be detectable from the outset of the session, and need not necessarily increase or decrease towards the end of the session. Thus the first and second halves should be more or less the same. With feedback, the scores might improve through a learning effect akin to biofeedback, and in fact there was a small but non-significant improvement (Table 1A). In trials without feedback, which are less interesting to take part in, there might be a tendency for subjects to get bored, or starers to concentrate less, leading to a decline. There was in fact a tendency for scores to decline, but it was not statistically significant (Table 1B).

In summary, the data from the first and second halves of these experimental sessions show a consistent pattern that goes against hypotheses that try to explain the results as artifacts. They support an interpretation in terms of the reality of the sense of being stared at.

20 Willow Road
London   NW3 1TJ

REFERENCES

Campbell, R. C. (1989) *Statistics for Biologists.* Cambridge: Cambridge University Press.

Carpenter, R. H. S.   (2005)   Does scopaesthesia imply extramission?   *Journal of Consciousness Studies 12,* 76–77.

Colwell, J., Schröder, S and Sladen, D.   (2000)   The ability to detect unseen staring: a literature review and empirical tests. *British Journal of Psychology 91,* 71–85.

Cottrell, J. E., Winer, G. A. and Smith, M. C. (1996) Beliefs of children and adults about feeling stares of unseen others. *Developmental Psychology 32,* 50–61.

Lobach, E. and Bierman, D. J. (2004) The invisible gaze: three attempts to replicate Sheldrake's staring effects *Proceedings of Presented Papers. The Parapsychological Association 47th Annual Convention.*

Marks, D. and Colwell, J.   (2000)   The psychic staring effect: an artifact of pseudo randomization. *Skeptical Inquirer (September/October),* 41–49.

Radin, D. (2004) The feeling of being stared at: an analysis and replication. *JSPR 68,* 245–252.

Sheldrake, R.   (1994)   *Seven Experiments that Could Change the World,*   Chapter 4. London: Fourth Estate.

Sheldrake, R. (1998) The sense of being stared at: experiments in schools. *JSPR 62,* 311–323.

Sheldrake, R.   (1999)   The 'sense of being stared at' confirmed by simple experiments. *Biology Forum 92,* 53–76.

Sheldrake, R.   (2000)   The 'sense of being stared at' does not depend on known sensory clues. *Biology Forum 93,* 209–224.

Sheldrake, R. (2001a) Experiments on the sense of being stared at: the elimination of possible artifacts. *JSPR 65,* 122–137.

Sheldrake, R. (2001b) Research on the sense of being stared at. *Skeptical Inquirer (March/April)*, 58–61.

Sheldrake, R. (2002) The sense of being stared at: an experiment at Holma College. *Gränsoverskridaren 10*, 21–23.

Sheldrake, R. (2003a) *The Sense of Being Stared At, And Other Aspects of the Extended Mind*. London: Hutchinson.

Sheldrake, R. (2003b) The need for open-minded scepticism: a reply to David Marks. *The Skeptic 16*, 8–13.

Sheldrake, R. (2005) The sense of being stared at. Part 1: Is it real or illusory? *Journal of Consciousness Studies 12*, 10–31.

Sheldrake, R., Overby, C. and Beeharee, A. (2008) The sense of being stared at: an automated test on the internet. *JSPR 72*, 86–97.

Wiseman, R. and Schlitz, M. J. (1997) Experimenter effects and the remote detection of staring. *JP 61*, 197–208.