THE INVISIBLE GAZE: THREE ATTEMPTS TO REPLICATE SHELDRAKE'S STARING EFFECTS

Eva Lobach, & Dick J. Bierman
University of Amsterdam
Amsterdam, The Netherlands

ABSTRACT

In three consecutive studies participants were asked to guess if someone was staring at them under conditions of sensory shielding. In total 188 sessions resulted in 4784 trials. The over-all hitrates were 50.6% (N=53), 52.1% (N=45), and 49.7% (N=37) resp. In study 1 both skeptics and believers acted as starer. Hitrate was marginally significant only for skeptic starers. Extraversion of the participant was irrelevant. In study 2 the over-all hitrate of 52.1% was not significant. In this study, skin conductance of the participant was also used as dependent variable. Initially, when participants were not informed of the staring, skin conductance showed no difference between stare and non-stare trials. After informing the participants, skin conductance was marginally higher for the stare trials. Study 3 compared hitrates for bonded pairs (friends) with non-bonded (strangers) pairs. In neither condition, not the slightest sign of a staring effect could be established. The conclusion of the three studies is that the staring paradigm is not the easily replicable paradigm that it is claimed to be. Apart from one internal effect all hitrates were very close to chance. Simulations with target sequences identical to those used in earlier studies involving feedback to the starees showed that results of these earlier studies can be explained by a combination of response bias and response strategies rather than psi. Other possible explanations of the failure to replicate concern difference in participants population, type of experimental setting, and rigor of the experimental procedures.

INTRODUCTION

Can we feel when someone is staring at us from behind? Some claim that we can.

One of the most ardent proponents of this claim is Sheldrake, who reports a large number of studies, almost all done with school children (ages 8 through 18) in school settings, showing an apparently repeatable, though as yet unexplained effect.

In a typical study, school children are paired up, one of them looks or does not look at the back of the blindfolded partner in a 20-trial sequence, after which roles of starer and staree are reversed. The starer indicates the start of a trial by a clicker, reponses of the staree (encouraged to be given quickly, often within ten seconds) are noted on a data sheet by the starer (Sheldrake, 2000b).

Responses at stare trials are consistently above chance, while they are around chance at non-stare trials (Sheldrake, 1998, 1999, 2000a, 2001a, 2003a). Sheldrake explains the different results for stare and non-stare trials as a natural phenomenon: we tend to be more aware when some stimulus is present than when it is absent.

A few attempts to replicate Sheldrake's results in more controlled laboratory settings have led to mixed results and to methodological controversies. One of these controversies concerns the possible learning effects of immediate feedback to the staree. According to Colwell, Schröder and Sladen (2000), feedback to starees will not improve their ability to detect an unseen gaze, but will only help to detect patterns in the sequences of stare and non-stare trials. They state that the sequences used by Sheldrake show more alternations (stare/non-stare) than would be expected if the sequences were truly random. When feedback is given, participants will be able to develop an anticipatory strategy that uses a similar alternating pattern as is apparent in the sequences used by Sheldrake. Also, feedback may encourage the staree to try to balance the number of yes and no responses. This strategy will increase the number of hits if the sequence contains an equal number of stare and non-stare trials, as was the case in many of Sheldrake's experiments. Both the alternating and the balancing response strategy may produce results that appear to show a staring effect while in fact no paranormal ability is involved. Sheldrake has since conducted additional studies without feedback, and the reported results are similar (Sheldrake, 2001a). Furthermore, new stare/non-stare sequences are currently in use that try to avoid a balanced number of stare and non-stare trials and the excess of alternating sequences.

One consistent point of critcism is the possibility of sensory 'leakage,' which may allow the staree to use the known senses to detect whether or not he is being stared at. Because the early experiments were done while starer and staree were sitting behind each other in the same room, sensory leakage could not be excluded. In response to this criticism Sheldrake has conducted or supervised a number of studies in which starers and starees are separated by glass windows with unknown sound isolation properties, and the reported results are still similar (Sheldrake, 2000a).

Response bias of the staree in favor of 'yes' ('I'm being looked at') has also been identified as a potential source of artefactual results. In Sheldrake's experiments participants respond 'yes' in about 55% of the trials. The greater number of hits at stare trials could thus have resulted from this response bias. However, response bias is considered to be a trait of the respondent, independent of condition. This means that hits due to response bias in the stare condition would be balanced out by misses in the non-stare condition. As results in the non-stare condition are at chance level, the over-all results cannot be attributed solely to response bias. Any statement about a *differential* sensitivity at stare *vs.* non-stare trials must be regarded skeptically, however, since these differences are very likely to be obscured by various effects, among them those of response bias and strategic guessing. In a formal description of staring experiments, Van Bolhuis (2004) explored the possibilities of testing the staring effect for stare and non-stare conditions separately under different sets of assumptions. He showed that this could only be done when sensitivity to staring is the sole factor influencing the responses. As soon as the model assumes the presence of any other influences, such as response bias, there is no way to tease apart their respective effects. On the positive side, it can be safely stated that an over-all hitrate is not sensitive to response bias. Since the over-all hitrates reported by Sheldrake are very consistently above chance, they must therefore be caused by 'something else.'

Sheldrake claims that this 'something else' involves a staring effect. The numerous experiments reported by Sheldrake supporting this claim suggest that this staring effect can also be easily replicated. This is rather remarkable, considering that many researchers often find it difficult to replicate psi phenomena, even more so when conscious measures are used as the dependent variable, as is the case in Sheldrake's studies. In the case of the staring effect, laboratory studies using skin conductance variations as the non-conscious dependent variable during stare and non-stare trials have shown less replicable results. Reviewing a number of these studies, the conclusion was that these studies show 'hints of an effect' (Schmidt, Schneider, Utts, & Walach, in press).

In light of the evidence reported by Sheldrake and his suggestion that these studies make great student projects (Sheldrake, 2003b), we decided to undertake attempts to replicate his findings. Because the three studies reported here were designed and conducted in collaboration with students fulfilling their second-year research project in psychology, they also contain some theoretical extras that reflect the students' interests. In Study 1, we took up the starer's belief in the staring effect and the staree's extraversion as independent variables, in Study 2 we measured skin conductance, and in Study 3 we compared starer-staree pairs of strangers and of friends and fed back the staree's responses to the starer. Common to all studies was the general experimental set up in which starer and staree were placed in separate rooms to avoid sensory leakage as much as we could. Also, no feedback was given to starees, and starees entered their responses directly into the computer.

Although we will report results separately for stare and non-stare trials, we will not conduct separate analyses on these data, because, as we pointed out above, these are heavily influenced by response bias. Instead we report statistical analyses of over-all hitrates only.

To allow for comparison with results reported by Sheldrake, we will also report the percentages of participants scoring more hits than misses, disregarding participants with an equal number of hits and misses. It should be noted, however, that this measure of hits and misses when split for stare and non-stare trials is also strongly affected by response bias, so we will again only test the over-all percentages of participants scoring more hits than misses. Unlike Sheldrake, we will use the binomial p-value, instead of using chi-square. The advantage of the binomial is that it allows for one-sided testing, which is appropriate because the staring effect hypothesis assumes that there will be more participants with hits than misses than vice versa.

If there is a staring effect, obviously there will be more 'yes' responses in the stare condition than in the non-stare condition. Besides reporting over-all response bias, we will also report response bias differences between stare and non-stare trials. We will refer to this difference in response bias as 'possibly psi-mediated response bias.'

STUDY 1

Besides attempting to replicate Sheldrake's results, Study 1 was also designed to explore the effects of the starer's belief in the staring phenomenon and whether the staree's introversion was related to successful staring detection.

Believers are reported to 'produce' stronger staring effects on skin conductance of the staree (e.g. Wiseman & Schlitz, 1997), and are in general found to produce stronger psi effects. The starer's belief was taken up as a within-subjects variable.

Skin conductance of introverts is reported to be more strongly affected by being stared at (Braud, Shafer & Andrews, 1993). We were interested whether introversion was also related to the number of correct responses when explicitly asked to indicate if they felt someone looking at them.

METHOD

Participants

Sixty four first year psychology students participated in exchange for course credit. Students with very high or very low extraversion scores on an earlier administered personality test (5PFT), were especially invited to participate. Others signed up on lists. All participants were informed of the purpose of the study.

Four participants were excluded because they did not cooperate fully, as was apparent from their behaviour, e.g. by reading a magazine during the experiment. Four more were excluded because they answered only 'yes' or only 'no' at all trials in one or both blocks and three more were excluded because they had more than 3 missing values (> 5%). This left 53 participants for data analysis, 39 women and 14 men.

Materials

- Rooms. Starer and staree were seated in adjacent rooms, separated by a one-way mirror. In contrast to Sheldrake's studies, participants in this study were seated in such a way that their left side was exposed to the starer's gaze through the one-way mirror. This way, we hoped to make the staring more interesting for the starers.

- Starers

Four starers, all women, took turns staring. Two of them believed that people can sense an unseen gaze, and two did not. Each participant was being stared at by a believer (believing starer condition) and a skeptic (skeptic starer condition) consecutively, during one block of 30 trials each, counterbalanced. During a stare trial, the starer looked at the participant without interruption and focused her thoughts on the participant. During a non-stare trial, the starer looked down and tried to think of something else.

- Randomization

A visual basic program generated random sequences consisting of an equal number of stare and non-stare trials. Each sequence was printed on a sheet of paper that was consulted by the starer in between trials. A different random sequence was used for each session.

- Trial synchronization

To synchronize trials for starer and staree, a taped voice indicated start and number of each trial on both sides of the oneway screen.

- Computerprogram

Trials lasted 15 seconds, with a 5 second intertrial interval. At the start of each trial the monitor showed the question "Do you feel that someone is looking at you now?" while a small clock in the corner of the screen counted backwards from 15 to 0 to indicate the time left to enter a response. The response could be given by hitting either one of two keys, meaning 'yes' or 'no'. If no answer was giving during this period, it was registered as missing. During the 5 second interval between trials, the computer presented the question "How certain were you of your answer?" Participants could indicate degree of certainty on a seven point scale ranging from 'completely uncertain' to 'completely certain'.

- Extraversion scale

The 14-item extraversion scale of the 5PFT, Cronbach's alpha = 0.85 (Elshout & Akkerman, 1999) measured introversion.

Procedure

One participant was invited at a time. Participants were told that during some of the trials someone would be looking at them from behind the one-way mirror, and that the order of stare or non-stare trials was random. They were instructed to concentrate and to base their answers on intuition and subtle differences in sensations or feeling. Participants did not meet the starer.

After three practice trials, the experimenter left the room and the first block of 30 trials started. After a 5 minute break a second block of 30 trials with another starer was performed. The procedure took between 30 and 40 min.

RESULTS

Of a total of 3180 trials, 36 (1.1 %) were missing because the participant did not respond in time. Hitrate, as reported in Table 1, is corrected for missing trials.

Table 1. Results of Study 1: Raw hits, hitrates, and percentages of participants with more hits than misses. N=53

CONDITION		# trials	**raw hits** M (*sd*)	**average hitrate**	**average over-all hitrate**	**percentage of participants with hits > misses**	**n[a]**	**over-all hits>misses**	**n[a]**
Believer	**Stare**	**15**	7.40 (*2.39*)	49.9%	48.9%	52.9%	51	44.0%	50
	Non-stare	**15**	7.19 (*2.24*)	48.5%		46.2%	52		
Skeptic	**Stare**	**15**	8.06 (*2.14*)	54.3%	52.3%	61.5%	52	58.7%	46
	Non-stare	**15**	7.51 (*2.27*)	50.6%		54.9%	51		
Total	**Stare**	**30**	15.45 (*3.53*)	52.1%	50.6%	59.6%	52	56.3%	48
	Non-stare	**30**	14.70 (*3.82*)	49.6%		43.1%	51		

[a] Participants with an equal number of hits and misses are disregarded

Average hitrates

Average hitrate (Table 1) was significantly different from the expected 50% only in the skeptic starer condition, (52.3%, t(52)=1.99, p=.03, one-tailed, not Bonferoni corrected), but not in the other condition, nor for the average total hitrates for the two conditions together.

Percentage of participants with more hits than misses

The percentage of participants scoring more hits than misses (Table 1) was never significantly higher than chance level. The skeptic starer condition came closest to statistical significance, exact binomial z = 1.03, p=.15, one-tailed.[1]

Response bias

Over-all response bias in favor of 'yes' ranged from 23% to 86%, average was 52%, not significantly different from the expected 50%, t(52)=1.05, p=.30, two-tailed.

Possible psi meditated response bias was not different in stare and non-stare conditions, except in the skeptic starer condition, M=54% at stare and M=50% at non-stare trials t(52)=1.92, p=.03, one-tailed.

Extraversion, hitrates and response bias

Mean Extraversion score was 62.9 (sd=13.8). We formed two groups of participants around the median. Over-all staring hitrates were the same for high- and low-extravert participants, as were the percentages scoring more hits than misses. On average, high-extravert participants responded 'yes' more often. This difference in over-all response bias was statistically significant only in the skeptic starer condition. Exploratory analysis showed that the above reported staring effect in the skeptic starer condition disappeared completely when controlling for extraversion. We also observed that the most succesful starer, who generated more hits than the other three in spite of being a skeptic, was the most extravert among the starers.

Discussion Study 1

Contrary to our expectation, the starees who are being looked at by a starer not believing in the effects of staring, appear to detect the staring more accurately than if they are being looked at by a believing starer. This result did not convert our skeptics, though, since it still concerns a very small effect of only 2.3%, not comparable to the larger effects reported by Sheldrake. The effect becomes non-significant (p = .12) after correction for multiple analyses. Note that the non-believing starers involved were no die-hard skeptics and were relatively open-minded about the phenomenon at hand.

Our results show no indication that introverts are better at detecting unseen staring. On the contrary, it seems as if the higher hitrates in the skeptic starer condition are mostly produced by the extraverted starees' greater tendency to say 'yes' when stared at by the skeptics, producing more hits at stare trials. Speculating on the observation that the most successful skeptic starer happened to be the most extraverted of the four starers, it seems possible that the extravered skeptic starer found it easier to bond with extraverted starees (or even vice versa) and that this might have had some effect in the desired direction. Another explanation could be that this extraverted starer happened to be very good at staring.

Similarities in personality between starer and staree and skill at staring might be an issue to explore in future studies.

STUDY 2

Considering the over-all rather disappointing results of Study 1, we wondered if our rigourous experimental set-up and the laboratory setting might have interfered with becoming consciously aware of

---

[1] Note that despite the much higher percentages than the overall hitrates, the analysis does not reach significance because much information has been discarded. Instead of an individual hitrate, a participant contributes only + or – to the analysis, which can thus only be tested using chi square (DAT BEGRIJP IK NIET, JE KUNT RUN SCORING GEWOON TESTEN MET EEN BINOMIAL TEST MET p=0.5 NET ZOALS JE MUNTEN WERPEN ZOU TESTEN) or binomial z.

subtle sensations as might be caused by being stared at. Psychophysiological measures, such as skin conductance, have on the other hand repeatedly shown to be affected by staring in a laboratory-setting (Braud & Schlitz, 1991; Schmidt et al., in press). Specifically, Braud and Schlitz (Braud & Schlitz, 1991) showed that skin conductance was lower during stare trials for participants and starers trained in feeling connectedness with each other and the world at large. In contrast, participants not so trained tended to have higher skin conductance when being stared at by a relative stranger.

In Study 2, we attempted to replicate the latter findings using a stranger as the starer, and using a Closed Circuit Television (CCTV) as employed in earlier experiments by Braud & Schlitz (1991). At the same time we wished to explore whether these presumably unconsciously produced physiological effects correlated with conscious responses.

Because informing participants of the purpose of the study might produce unwanted side-effects on skin conductance, the session was split in two blocks. In the first block (naive condition), participants were told that they were to judge the 'tranquility' of musical fragments, while we simultaneously measured their skin-conductance. During half of the trials, unbeknownst to the participants, they would be stared at through a hidden camera connected to a monitor in an adjacent room. In the second block (informed condition), participants were shown the camera and told that they were being observed during some of the trials. In this second block, instead of judging musical fragments, they had to answer the question whether they felt they were being looked at.

We expected to find higher skin conductances during stare trials in both the naive and the informed block with an effect size larger than the one obtained by conscious guessing. In addition, we were interested whether variations in skin conductance correlated with correct guesses. Also, if the participants' answers ('yes' or 'no') correlated with skin conductance variations, irrespective of stare or non-stare condition, this suggests that their answers might be prompted by these variations.

METHOD

Participants

59 first year psychology students signed up on a list indicating the purpose of the study as 'judging musical fragments', and participated in exchange for course-credit.

Ten participants were excluded because of malfunctioning of GSR-equipment. Four more were excluded because participants answered 'no' at all trials.

This left 45 participants for data analysis, 15 men and 30 women.

Materials

- Rooms. Staree and starer were in separate but adjacent rooms. The experimental room contained a relaxing chair in front of a monitor, the equipment to measure skin conductance, and a response box on which the participants could indicate their answers. A hidden videocamera, behind a binder on a shelf, relayed an image of face and shoulders of the participant in black and white to a monitor in the adjacent observation room. The starer in the observation room sat at the desk that supported the equipment registering the GSR, a monitor showing the screen presented to the participant so the starer would know the number, and start and end of the trial, and a monitor displaying the image of the staree. This monitor was covered by a heavy cloth during the whole procedure, lifted by the starer during stare trials only.
- Trials. In both blocks trials lasted 30 s with a 10 s interval in which the participant answered the question presented on the monitor.
- Musical fragments. 20 different fragments of soft meditation music were presented through two loudspeakers in trials of 30 seconds each.
- Starers. All five experimenters, four women and one man, took turns staring. The starer never left the room and never met the participant.
- Randomization. Twenty identical slips of paper, ten with 'stare' and ten with 'non-stare' written on them were shuffled thoroughly and then were picked one by one, noting down their order. This procedure

was repeated ten times to create ten different random orders, which were printed out on ten identical sheets of paper. Before each block, the starer picked one of the ten sheets blindly. Before each pick, the sheets were shuffled. If the starer picked the same sequence for the second block, the procedure was repeated.

- Computer program. During the first part of the session (naive condition), the computer program indicated the start of the trial by displaying the number of the trial on the staree's monitor and presenting one of the twenty 30 s musical fragments. Each staree heard all twenty musical fragments. During each 10 s inter trial interval, the monitor presented the question: "Did you find this music relaxing?" Participants indicated their response 'yes' or 'no' by pressing a button on the response box with their dominant hand.

The same general procedure was used during the second session (informed condition) but instead the monitor asked: "Did you feel you were being looked at?"

Procedure

The experimenter attached the Ag-AgCl electrodes of the UvA GSR-equipment to the participant's index and middle finger of their non-dominant hand while explaining the procedure. The participant was then left alone for a 10 min acclimatization period after which the experimenter checked if everything was all right and started the experimental program.

After twenty trials the experimenter returned to give instructions for the second block. The whole procedure took about 50 minutes.

RESULTS

Average hitrate

The average hitrate of 52.1 % (Table 2) is higher than chance, but not significantly so, t(44)=1.24, p=.11, one-tailed. Note that the apparent differential effect between stare and non-stare trials can not be disentangled from response bias.

Percentage of participants scoring more hits than misses

Binomial z testing the percentage of participants scoring more hits than misses (Table 2) showed that it was not above chance, z=.50, ns. As with the average hitrate, analysing results of stare and non-stare condition separately would again have given rise to misleading conclusions.

Table 2. Results of Study 2: Raw hits, hitrates, and percentages of participants with more hits than misses. N=45

CONDITION	# trials	**raw hits** M (*sd*)	**average hitrate**	**average over-all hitrate**	**hits > misses**	**n**	**over-all hits>misses**	**n**
Stare	**10**	4.38 (*2.01*)	43.8%	52.1%	42.9%	35	55.6%	36
Non-stare	**10**	6.04 (*2.28*)	60.4%		73.5%	34		

Response bias

Over-all response bias in favor of yes ranged from 10% to 80%. The average was 42%, significantly different from the expected 50%, t(44) = - 3.07, p < .005 (two-tailed). As in the skeptic condition in Study 1, possibly psi-mediated response bias was higher at stare (44%) than at non-stare trials (40%), but this time the difference was not statistically significant, t(44)=1.24, p=.11 (one-tailed).

Skin conductance

In the first, naive, block, mean skin conductance did not differ in stare (M=18.71, sd=5.79) *vs.* non-stare trials (M=18.73, sd=5.80) as tested in a paired samples t-test, t(44) = -.61, *ns*. In the second, informed, block however, participants showed a marginally higher skin conductance in stare (M=19.26, sd=5.85) *vs.* non-stare trials (M=19.15, sd=5.99), t(44) = 1.64, p = .054 (one-tailed).

Skin conductance and conscious verbal response

Skin conductance when the participant said 'yes' as compared to 'no' was on average slightly higher. The average difference was about 1 %, M= .14 (sd=.41), t (44)=2.15, p = .04 (two-tailed), suggesting that the response to say 'yes' is associated with higher skin conductance.

There was no indication that differences in gsr between stare and non-stare trials were related to the number of correct guesses.

Discussion Study 2

When participants were unaware that they were being looked at, their skin conductance was on average not different between stare or non-stare trials. When they were informed, however, skin conductance did marginally show the expected higher values at stare trials. Assuming that the skin responds to unseen gaze without conscious awareness, it is surprising that skin conductance only showed a difference when participants knew about the hidden camera. An explanation might be that knowing about the staring could have made the participant's attention more attuned to subtle variations, resulting in a stronger skin response. Another explanation is that the soothing music in the naive condition, which by necessity always preceded the informed condition, induced a mood that rendered the participant more open to being influenced by the staring. Both ideas could be experimentally tested by varying knowledge about the staring or by letting the actual experiment be preceded by relaxing music or not.

STUDY 3

For Study 3 we communicated elaborately with Sheldrake in order to remove any reasons for our failure to replicate. We were intrigued by Sheldrake's explanation that sensitivity to an unseen gaze might have survival value: animals sensing the predator's eyes have more chances to escape and thereby to transmit this sensitivity to their offspring (Sheldrake, 2000c) Although generally people do not relate to each other as predator and prey, we would expect the gaze of a stranger to be more threatening than the gaze of a more familiar person, as was supported by the findings that strangers tend to increase skin conductivity (Braud & Schlitz, 1991), and for which we also found some support in Study 2. In other contexts, Sheldrake suggests that a close relationship or a family bond is particularly advantageous for a staring effect to occur, as was repeatedly shown in experimental findings reported by Sheldrake (e.g. Sheldrake, 2001a). This is also in line with the tentative suggestion from observations made in Study 1, that similarities in personality between starer and staree might create a better bonding and subsequently more hits.

As a first attempt to create more clarity in this matter, we wished to compare success at staring between friends and between strangers. We invited pairs of people with a strong positive bond into our laboratory and asked them to act once as starer and staree together with their partner, and once again as starer and staree with a stranger who was a member of another pair that was invited at the same time. Again, a oneway mirror separated starer and staree, and the starees responded using the keyboard of a computer. During the duration of each trial (10 s) the monitor displayed the image of a flower.

The procedure is similar to the procedure used by Sheldrake in the sense that normal participants acted both as starer and staree. To make participating more interesting in the staring role, we decided to provide feedback on the responses and on the over-all hitrate of the staree. This allows the starer to check the effectiveness of various staring techniques (*cf*. Braud & Schlitz, 1991).

METHOD

Participants

Forty-four participants were recruited in pairs of good friends or relatives among family and friends of the experimenters in exchange for 7 Euro. Two pairs participated at the same time in a four-part session. All participants took turns as starer and staree, once with the partner they signed up with (*friends* condition), and once with a member of the other pair (*stranger* condition), counterbalanced.

Because one of the participants became unwell during a session, the data of the four participants in that session were lost. Three more were excluded because they only answered 'no' in one or both blocks. This left data of 37 participants for analysis, 10 men and 27 women.

Method

- Rooms. Two adjacent oneway-mirror units were used simultaneously. Starer and staree were both seated in front of a monitor. Both monitors displayed the same image at all times, but the bottom part of the screen, showing instructions to the starer, was hidden from the staree behind a strip of paper taped to the monitor. The starer's monitor was facing perpendicular to the one-way screen in such a way that the starer would only have to glance sideways to check whether it was a stare or non-stare trial, and no movement of the chair was necessary to adjust posture from a stare to a non-stare trial or vice versa.

- Computerprogram. To synchronize the procedure at both sides of the screen, the start of each 10s trial was indicated by a loud beep that was also audible on the other side of the one-way mirror. At the beep a picture of a flower was presented and stayed on screen for the duration of the trial, followed by the request: "Give your impression if you are being looked at now," together with a representation of the four response options "Yes I feel I am being looked at and I am certain," "Yes I feel I am being looked at and I am uncertain," "No, I don't feel I am being looked at and I am uncertain," and "No, I don't feel I am being looked at and I am certain." The staree indicated the appropriate answer by pressing any of four keys representing the four options during the 10 s intertrial interval.

At the bottom of the starer's monitor the type of trial (stare or non-stare) was indicated directly at the beep indicating the start of the trial. It was replaced by feedback about the response of the staree and cumulative number of hits during the intertrial interval.

- Randomization. The computer presented twenty[2] trials (ten stare and ten non-stare) in a random order, different for each participant.

Procedure

Two pairs of participants were invited at a time. Each pair was assigned to one of the two experimenters to guide them to their rooms and give instructions throughout the session. A session consisted of four parts. Each participant acted as starer and staree, once together with a friend (part one and two) and once together with a stranger (part three and four). In half of the sessions pairs started with the stranger condition; pairs were split up and coupled to a member of the other pair. In the other half of the sessions, pairs first worked together as starer and staree. A toss of a coin determined who would be starer first.

The staree was seated in front of a monitor, his back to the one-way mirror. The starer was seated behind the mirror. The participants were given print-outs of the instructions and allowed some time to read. Besides explaining the procedure, the instructions advised both starer and staree to focus on the image on the screen and to relax. After three practice trials the experimenters left the participants. During the second part of the session, the experimenters let starer and staree switch places and handed them instructions on their new roles. During the third and fourth part of the session, new pairs were formed. At the end of the session, participants filled out a short questionnaire asking them about their belief in the

[2] Ten more trials were presented in a fixed order as the middle ten of a total of 30 trials to test another hypothesis that we will not discuss here. They are left out of the analyses presented here

phenomenon, about their current feelings towards the friend they had signed up with, and whether they knew the member of the other pair they were paired up with for the stranger condition.

RESULTS

Hitrate and percentage of participants scoring more hits than misses

In none of the conditions was the hitrate higher than chance level, and neither was the percentage of participants scoring more hits than misses. (Table 3)

Table 3. Results of Study 3. Raw hits, hitrates, and percentages of participants with more hits than misses. N=37

CONDITION		# trials	**raw hits** M (*sd*)	**average hitrate**	**average over-all hitrate**	**hits > misses**	**n**	**over-all hits>misses**	**n**
Friends	**Stare**	**10**	5.00 (*1.58*)	50.0%	48.9%	42.4%	33	47.1%	30
	Non-stare	**10**	4.78 (*1.65*)	47.8%		50.0%	28		
Strangers	**Stare**	**10**	4.73 (*1.54*)	47.3%	50.4%	32.1%	28	53.3%	33
	Non-stare	**10**	5.35 (*1.70*)	53.5%		55.6%	27		
Total	**Stare**	**20**	9.73 (*2.64*)	48.7%	49.7%	40.0%	35	47.1%	34
	Non-stare	**20**	10.14 (*2.64*)	50.7%		40.0%	30		

Response bias

Response bias in favor of yes ranged from 25% to 70%, average was 49%, not significantly different from the expected 50%, t(36)= -0.36, *ns*. There was no indication of a possibly psi-mediated response bias because in none of the conditions did response bias differ between stare and non-stare trials. However, participants gave on average more 'yes' reponses in the friends condition (M=51%) than in the stranger condition (M=47%), a statistically significant difference, t(36)= 2.19, p=.04 (two-tailed).

Discussion Study 3

Results in Study 3 not only failed to establish whether the gaze of a friend can be detected more easily than the gaze of a stranger, it also failed to produce any indication that people are sensitive to being stared at from behind. One possible explanation for this negative result is a somewhat unfortunate aspect in the design that we only realised while past midway conducting the experiment: all participants saw all images of flowers in exactly the same order for all four parts of a session. When the participant acted as starer, in either the first or the second part of a session, this could easily create an association between a particular image and the type of trial, stare or non-stare, which might prompt an answer during the next part when the participant acted as staree. Assuming this is the case, only the first part of each session would be free of this kind of association. Testing hitrates from these first parts of the sessions, however, did still not show any indication of a staring effect.

Another possible artifact was the position of the starer behind the one-way screen. Because the experimenters were not present to check if starers followed strictly protocol, it is possible that their position allowed sideways glances through the screen. Although this could have happened, it seems rather unlikely that this would have happened very often or out of carelessness or lack of commitment, since almost all participants were good friends or close relatives of the experimenters and were in general highly motivated to collaborate and follow instructions carefully. Nevertheless, accidental sideways glances could have 'contaminated' the non-stare trials and thereby diluted a potential staring effect.

GENERAL DISCUSSION

The three studies presented above show a rather disappointing picture as far as the replicability of the staring effect is concerned. None of the studies found the large staring effects reported by Sheldrake. The straightforward conclusion must be that staring effects are not easily replicable, although they do indeed serve as great student projects.

We are left with the question of how to explain the discrepancy between Sheldrake's results and ours. Several answers spring to mind.

An important difference was the use of feedback in many of Sheldrake's earlier staring studies. It seems possible that a combination of response bias in favor of yes and guessing strategies in response to feedback could explain his results. In order to test this possibility, we ran a simulation using the stare/non-stare sequences as published at Sheldrake's website, http://www.sheldrake.org. There are two kinds of target sequences. The old 24 target sequences were created in response to recommendations by Wiseman and Smith (1994). These old sequences contained an equal number of stare and non-stare trials, and 12 sequences were the mirror image of the other 12. Criticisms by Colwell et al. (2000) about the large number of alternations prompted the creation of 20 new sequences that are all different, do not have a balanced number of stare and non-stare trials and do not show as many alternations as the old sequences.

We ran three different simulations on both the old and the new series of sequences, using 100 artificial subjects for each target sequence. The first was a mere chance simulation, without response bias and without guessing strategy, the second used a response bias of 55% in favor of 'yes', and the third used a response bias of 55% together with a guessing strategy in response to feedback.

The guessing strategy assumed that the participant adjusts his guessing probabilities ('yes' and 'no') to the feedback on the previous trials. This guessing strategy, which is akin to a gambler's fallacy, can be expressed (without response bias) as the probability to say 'yes' after feedback of n_stare + n_nostare trials:

p_say_yes_WB = (10 - n_stare) / (20 - n_stare - n_nostare)

where p_say_yes_WB is the probability the participant says 'yes' (without response bias), n_stare is the number of times that according to the feedback the subject was stared at and n_nostare the number of times the subject was not stared at. At the beginning of the session no information is available, n_stare and n_nostare are of course zero, and the probability to say 'yes I am being looked at' is 0.5.

For the third simulation this guessing strategy was combined with a response bias of 55% in favor of 'yes'. The strategy used in this simulation can be expressed as:

p_say_yes_B = p_call_looking_WB + bias

Table 4. Hitrates in a simulation with Sheldrake's 24 old and 20 new target sequences with No Feedback & No Response Bias, with No Feedback & Response Bias 55% and with Feedback & Response Bias 55%.

Simulation condition		Old target sequences (n=2400)	New target sequences (n=2000)
		hitrate	**hitrate**
No feedback – No bias	**Stare**	50.7%	49.9%
	Non-stare	49.6%	48.6%
No feedback -- Bias 55%	**Stare**	54.9%	55.1%
	Non-stare	45.1%	43.8%
Feedback[a] – Bias 55%	**Stare**	59.1%	59.0%
	Non-stare	49.5%	44.7%

[a] Simulation with feedback assumes that participants use a gambler-fallacy type of strategy. See text.

The results of these simple simulations are given in Table 4. We can see that the results of the mere chance simulation, without strategy and without response bias, look quite normal. This is true for the hitrates in both stare and non-stare condition. The simulation with a response bias of 55% shows the expected above chance hitrate in the stare condition, which is compensated by about the same under chance scoring so that the over-all results are at chance.

The most interesting results are those of the third simulation, using the above guessing strategy in combination with a response bias. The results thus obtained for the old (balanced) sequences mimic quite closely the results reported in the literature: they are at chance in the non-stare condition while those in the stare condition are around 60%. This of course can be expected given the balanced nature of the old target sequences. To our surprise, however, the very same strategy also yielded artificial psi scoring for the new sequences. That is, if one pools both conditions there is still a clear above chance scoring percentage of about 51.8%, statistically extremely significant. This unexpected result suggests that even among the new target sequences there are still too many that are close to a balanced sequence. On inspection, 16 of the 20 new target sequences have a 10-10 or a 9-11 distribution, and only 4 have a 8-12 distribution.

These simulations show that the effects found in staring studies involving feedback to the staree might explain the difference between the effects reported by Sheldrake and the absence of those effects in ours.

However, as we noted above, Sheldrake has since reported quite a number of studies that do not provide feedback to the staree, and these show somewhat smaller, but still large effects (Sheldrake, 2001b), unlike our three studies, so it leaves our question in part still unanswered.

Another important difference is that Sheldrake's experiments are almost exclusively conducted with school children from age 7 or 8 and up, while our experiments were mostly conducted with young adults, or older. However, Sheldrake has reported a number of studies with adults as well, with similar results.

Another difference is the use of blindfolds. We never used blindfolds in our studies, while Sheldrake has adopted the use of these in many of his staring experiments. It seems possible that blindfolds help participants to focus on the presumably subtle sensations involved in staring.

Still another difference is the experimental setting: the familiar environment of the school *vs*. a relatively sterile laboratory setting. Although the familiar school environment might provide the relaxed atmosphere that promotes psi phenomena, it also makes it much harder to adhere to controlled circumstances, which points to still another difference between our experiments and those of Sheldrake. In the latter the scoring is done by hand, and there are no double scoring sheets which would allow for double checking. One scoring error per sheet in favor of the staring hypothesis would suffice to get the results that Sheldrake reports. Although it is extremely unlikely that so many scoring errors have been made, adopting either independent hand scoring or automatic registration with full date and time information would prevent this concern. This way, the best of both approaches could be combined. The results of the laboratory studies that we conducted so far, leave no other conclusion than that the conscious effects of an unseen gaze remain invisible.

References

Braud & Schlitz (1991). Consciousness interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies, 2-1,* 1-46

Braud W., Shafer D., & Andrews S. (1993). Further studies of autonomic detection of remote staring: replication, new control procedures, and personality correlates. *The Journal of Parapsychology, 57,* 391-409.

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

Elshout, J.J., & Akkerman, A.E. (1975). *Vijf persoonlijkheidsfaktoren test, 5PFT [Five personality factors test, 5PFT],* Berkhout, B.V. – Nijmegen, The Netherlands.

Schmidt, S., Schneider, R., Utts, J. M. & Walach, H. (in press). Distant Intentionality and the Feeling of Being Stared At - Two Meta-Analyses. *British Journal of Psychology*.

Sheldrake R. (1998). The Sense of Being Stared At: Experiments in Schools. *Journal of the Society for Psychical Research, 62,* 311-323.

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

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

Sheldrake, R. (2000b) Research on the sense of being stared at. Available: http://www.sheldrake.org/experiments/staring/staring_experiment.html

Sheldrake, R. (2000c). *Dogs that know when their owners are coming home and other unexplained powers of animals.* Arrow Books – London, United Kingdom.

Sheldrake, R. (2001a). Experiments on the sense of being stared at: The elimination of possible artefacts. *Journal of the Society for Psychical Research, Vol. 65,* 122-137

Sheldrake, R. (2001b). Follow-up. Research on the feeling of being stared at. *Skeptical Inquirer magazine, march/april* Available: http://www.scicop.org/si/2001-03/stare.html

Sheldrake, R. (2003a). *The sense of being stared at and other aspects of the extended mind*, Hutchinson – London, United Kingdom.

Sheldrake, R. (2003b). Lecture given as part of the Sentient Creatures Series at the Waag, Amsterdam, The Netherlands, 8 October 2003. Available: http://connected.waag.org/archive.html

Van Bolhuis, J. (2004). An analysis of response bias in some of Rupert Sheldrake's experiments. *Unpublished document,* Free University, Amsterdam, The Netherlands.

Wiseman, R. & Schlitz, M. (1997). Experimenter effects and remote detection of staring. *Journal of Parapsychology, 61*, 197-208.

Wiseman, R. & Smith, M. (1994). A further look at the detection of unseen gaze. *Proceedings of the Parapsychological Association 37th Annual Convention.* Parapsychological Association. 465-478.