strong justification for further and better quality investigations of this intriguing topic in the future.

References

Braud, W., Shafer, D. & Andrews, S. (1993), 'Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection', *Journal of Parapsychology*, **57**, pp. 373–90.

Schmidt, S., Schneider, R., Utts, J. & Walach, H. (2004), 'Distant intentionality and the feeling of being stared at: Two meta-analyses', *British Journal of Psychology*, **95**, pp. 235–47.

DEAN RADIN

The Sense of Being Stared At: A Preliminary Meta-Analysis

Abstract: *Meta-analysis of 60 experiments investigating the conscious sense of being stared at suggests that the reported effects may reflect a genuine ability. A subset of 10 of these studies, designed to preclude implicit learning of sensory cues, resulted in a homogeneous distribution of effect sizes and a weighted mean effect size substantially beyond chance expectation ($p = 5 \times 10^{-17}$).*

Two types of experiments have been conducted to investigate the commonly reported 'sense of being stared at' — those based on conscious reports and those based on unconscious physiological responses. The first class, described by Sheldrake in this issue, studies the ability of a 'staree' to consciously detect being stared at, typically by a 'starer' located behind the staree. This paper reports a preliminary meta-analytic examination of these experiments.

The second class of experiments, reported by Schlitz and Braud (1997) and others, investigates whether a starer's gaze over a closed-circuit video link, under conditions that rigorously exclude sensory cues, can be detected as unconscious fluctuations in a staree's skin conductance (Braud & Schlitz, 1989; 1991). Schmidt *et al*. (2004) report a meta-analysis of 15 such experiments involving a total of 379 individual testing sessions. That analysis provided significant evidence for a distant staring effect, a homogeneous distribution of effect sizes, and no evidence of a selective reporting problem. Those studies suggest that an unconscious sense of being stared at is a genuine, independently repeatable effect, providing proof-of-principle for the effects reported by Sheldrake.

Method

I excluded from consideration a large-scale staring study conducted at the NEMO Science Centre in Amsterdam and also web-based staring studies, as trials collected in those studies were unsupervised (Sheldrake, this issue). From Sheldrake (1998; 1999; 2000a,b; 2003), and citations therein (e.g., Coover, 1913; Colwell *et al*., 2000; Lobach & Bierman, 2004; Poortman, 1959; Radin, 2004), I identified 60 relevant experiments reporting a total of 33,357 trials collected under one or more investigator's supervision. These studies involved three categories of control for implicit learning of sensory cues: tests conducted in close proximity with trial-by-trial feedback, tests conducted in close proximity without feedback, and tests conducted by looking through windows and without feedback. Of the 60 studies, the majority (88%) fell in the first and third categories.

The raw data in each study was the number of times the staree successfully identified being stared at or not stared at (each success called a *hit*), and the total number of trials conducted in the study (N). Hits and trials were used to form a per study 'hit rate' as $p_1 = hits/N$, and then p_1 was used to form a standard normal variable $z = (p_1 - p_0)/se$, where $p_0 = 0.5$ (meaning an equal *a priori* likelihood of being in a staring or non-staring condition), $se = \sqrt{p_0 q_0 / N}$ and $q_0 = 1 - p_0$. Effect size per trial (per study) was determined as $e = z/\sqrt{N}$, and a weighted mean effect size assuming a fixed effects model (FEM) was formed with a weighting factor based on the inverse of the squared standard error, $w = 1/se^2$, which in this case is equivalent to $w = 4N$ (Hedges, 1994). Then a weighted mean effect size was formed as $\bar{e} = \sum (w \times e) / \sum w$, where the sums are taken over N studies. To assess the homogeneity of effect sizes, the statistic Q was determined, where $Q = \sum (w \times e^2) - \left[\sum (w \times e)\right]^2 / \sum w$ with N–1 degrees of freedom. Q follows the chi-squared distribution.

Results

Figure 1 shows the cumulative mean hit rate across the 60 studies, with one standard error bars. It is clear that the mean hit rate stabilizes to just over 54%, where 50% is expected by chance. The FEM weighted mean effect size was significantly above chance, $\bar{e} = 0.089 \pm 0.003$ (mean ± standard error), $z = 32.5$, $p = 10^{-232}$, and the distribution of effects was significantly heterogeneous, $Q = 763.3$ (58 df), $p = 10^{-123}$. Because of the heterogeneity, a more conservative random effects model was also determined (REM, Hedges & Vevea, 1998). The

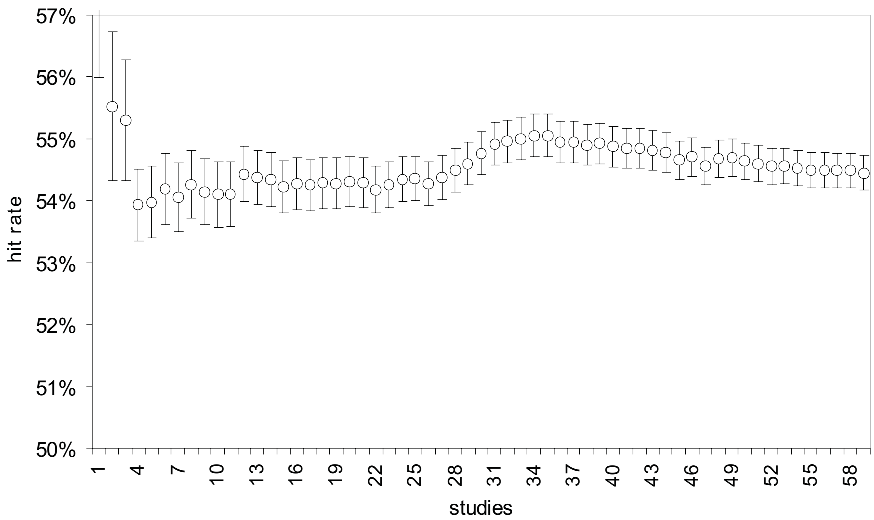

Figure 1. Cumulative mean hit rate for 60 sense-of-being-stared-at experiments, with one standard error bars.

REM weighted mean effect size remained significantly above chance, $\bar{e} = 0.114 \pm 0.010$, $z = 10.9$, $p = 10^{-28}$, and the distribution of effects also remained heterogeneous, Q = 167.8 (58 df), $p = 10^{-12}$.

Discussion

One possible explanation for the heterogeneity is a selective reporting bias. A common way to visualize whether selective reporting might be a problem is with a 'funnel plot' graphing effect size vs. sample size. If such a plot is positively skewed, it indicates that small sample-size studies with negative outcomes were probably not reported. Use of a technique known as 'trim and fill' can be applied to this plot to algorithmically identify and fill in the potentially missing studies to make the plot symmetric (Duval & Tweedie, 2000). A new weighted mean effect size can then be formed using the missing studies to form a conservative estimate of the 'true' effect size.

The black circles and white diamond in the funnel plot in Figure 2 show the effect size for each of the 60 reported studies. The positive skew in this plot indicates that this database probably has a selective reporting problem. The trim and fill algorithm identified that six studies were required to make the plot symmetric; these are shown as the white circles in Figure 2. The dashed vertical line indicates the FEM weighted mean effect size. Recalculating both the FEM and REM means effects after adding the six estimated studies resulted in slightly smaller but still highly significant mean effect sizes: $\bar{e}$ (FEM) = 0.078 ± 0.003, $z = 28.9$, $p = 10^{-184}$ and $\bar{e}$ (REM) = 0.072 ± 0.010, $z = 7.15$, $p = 10^{-13}$.

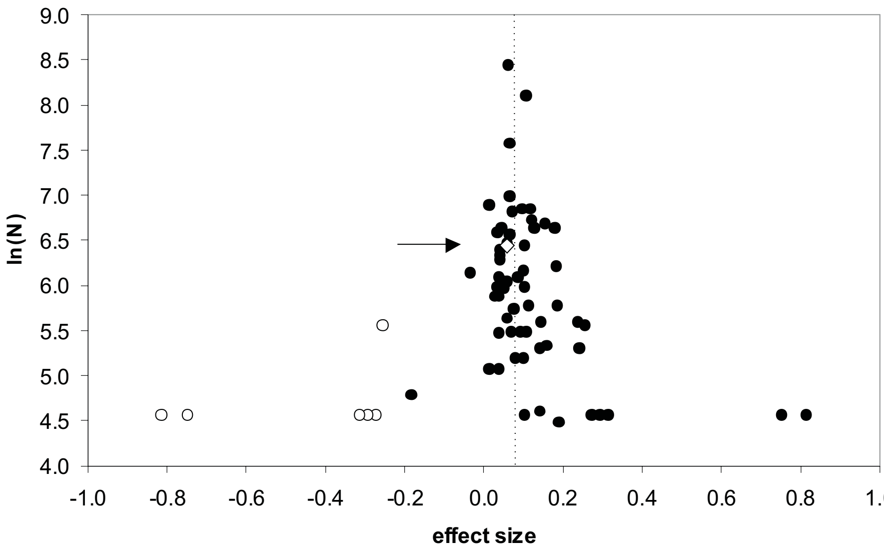

Figure 2. Funnel plot for 59 sense-of-being-stared-at experiments (black dots) and six studies identified by the 'trim and fill' algorithm (white dots). The arrow pointing to the white diamond indicates the effect size measured in a recent replication attempt using a computer-based, automated recording system (Radin, 2004).

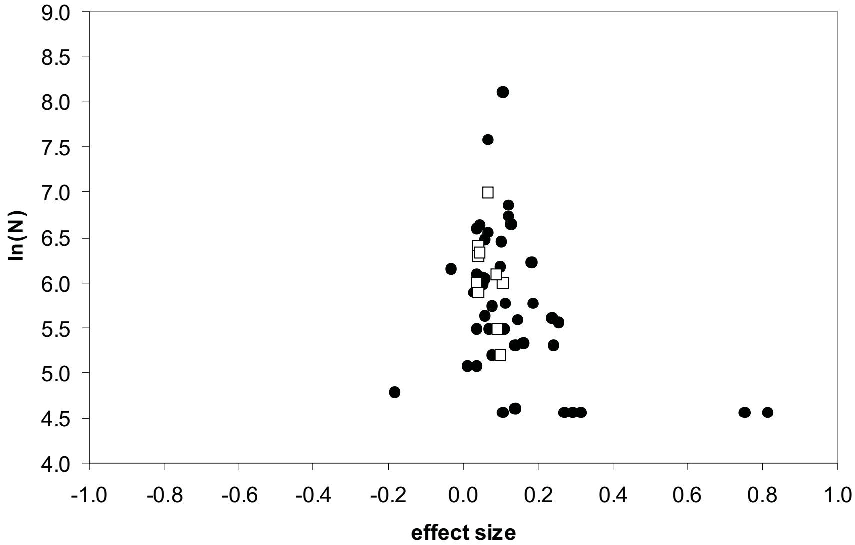

Figure 3. Funnel plot comparing 42 experiments conducted in close proximity and with trial by trial feedback (black circles), and 10 experiments conducted at a distance, looking through a window, and without feedback (white squares).

By eliminating 19 studies with the largest Q values, the effect size heterogeneity becomes homogeneous. All of the eliminated studies involved close proximity designs with trial-by-trial feedback. With the remaining homogeneously distributed studies, the FEM and REM mean weighted effect sizes remain significant; the more conservative e (REM) = 0.063 ± 0.007, $p = 10^{-19}$.

To examine the possibility that studies with especially large outcomes might have been due to poorer controls for sensory cues, two funnel plots were formed, as shown in Figure 3. The black circles in this figure show 42 studies involving designs with close proximity and trial-by-trial feedback, and the white squares show 10 studies conducted through windows without feedback. It is evident that both the heterogeneity and the selective reporting problem are due to the close-proximity studies. The 10 better-controlled studies are homogenously distributed and do not show evidence of selective reporting. The FEM weighted mean effect size for the through-the-window studies (which is the appropriate effect size model given the homogeneity of effect sizes) is highly significant ($p = 4.8 \times 10^{-17}$). Comparison of the FEM and REM models is shown in Table 1.

Comparison of the close-proximity vs. the through-the-window studies shows that the latter had significantly lower mean effect sizes in both fixed and random effects models ($z = -4.96$, $p = 10^{-6}$, two-tailed; $z = -2.00$, $p = 0.05$, two-tailed, respectively). This suggests that the mean effect size for close-proximity studies might be inflated due to implicit learning of subliminal cues, but that explanation is confounded by the likelihood of selective reporting, as shown by the funnel plot and suggested by the heterogeneity of this class of experiments.

	Through window		Close proximity	
Model	FEM	REM	FEM	REM
e	0.060	0.064	0.100	0.131
se	0.007	0.030	0.004	0.015
z	8.31	2.14	27.84	8.78
p	4.8×10^{-17}	0.016	6.4×10^{-171}	8.3×10^{-19}
Q	10.74	0.80	648.75	115.40
df	9	9	41	41
p	0.29	0.99	2.4×10^{-110}	5.3×10^{-9}

Table 1. Comparison of fixed and random effects models for sense of being stared at studies conducted with higher controls for implicit learning of sensory cues (staring through windows without feedback) and lower controls (staring at close proximity with feedback).

One way to assess the impact of the selective reporting on the overall evidence is to estimate how many missing studies are required to nullify the overall effect. Two models of estimating this missing 'filedrawer' were employed. The first, proposed by Rosenthal (1979; 1991), asks how many studies averaging an overall null effect would be required to reduce the overall level of significance to just above the conventional level of statistical significance, i.e. $p > 0.05$. The second, proposed by Scargle (2000) and independently by Hsu (2002), asks the same question but assumes that the missing studies would average to a slightly negative effect rather than to a null effect. The Scargle/Hsu assumption is based on the expectation that all studies with positive outcomes at $p < 0.05$ (one-tailed) *or* with negative outcomes at $p < 0.05$ (two-tailed) are published. The asymmetry in thresholds for assigning statistical 'significance' to positive and negative outcomes provides a more conservative filedrawer estimate.

Table 2 shows that the estimated filedrawer contains between 1,417 and 7,729 missing studies, depending on which model one wishes to adopt. Converting these figures into an estimated number of missing individual trials, based on the average number of trials per reported study ($N = 565$), results in the range of 800,000 to 3 million missing trials. These figures seem implausible given the relatively small numbers of investigators who have demonstrated interest in this line of experimentation.

Method	Filedrawer estimate	Trials
Rosenthal	7,729	3,045,210
Scargle/Hsu	1,417	800,885

Table 2. Filedrawer estimates using Rosenthal and Scargle/Hsu methods.

As more of these studies are conducted, new meta-analyses will undoubtedly assess in finer detail design elements such as the types of controls employed, how conditions were randomized, how the tests were supervised, how data were recorded, and so on. Examination of such moderator variables will help explain the overall heterogeneity. Based on Schmidt *et al*.'s (2004) meta-analysis of similar experiments involving rigorous controls against sensory cues, it seems likely that future meta-analyses will continue to show that some proportion of these effects reflect a genuine but poorly understood 'sense'.

References

Braud, W. and Schlitz, M. (1989), 'Possible role of intuitive data sorting in electrodermal biological psychokinesis (bio-PK)', *Journal of the American Society for Psychical Research*, **83** (4), pp. 289–302.

Braud, W. and Schlitz, M. (1991), 'Consciousness interactions with remote biological systems: anomalous intentionality effects', *Subtle Energies and Energy Medicine*, **2**, pp. 1–46.

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

Coover, J.E. (1913), 'The feeling of being stared at – experimental', *American Journal of Psychology*, **24**, pp. 570–5.

Duval, S. & Tweedie, R. (2000), 'A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis', *Journal of the American Statistical Association*, **95**, pp. 89–98.

Hedges, L.V. (1994), 'Fixed effect models', in *The Handbook of Research Synthesis*, ed. L.V. Hedges & H. Cooper (New York: Russell Sage Foundation).

Hedges, L.V., & Vevea, J.L. (1998), 'Fixed- and random-effects models in meta-analysis', *Psychological Methods*, **3**, pp. 486–504.

Hsu, L.M. (2002), 'Fail-safe Ns for one- versus two-tailed tests lead to different conclusions about publication bias', *Understanding Statistics*, **1** (2), p. 85–100.

Lobach, E. & Bierman, D.J. (2004), 'The invisible gaze: Three attempts to replicate Sheldrake's staring effects', *Proceedings of the Parapsychological Association Annual Convention*, 2004.

Poortman, J.J. (1959), 'The feeling of being stared at', *Journal of the Society for Psychical Research*, **40**, pp. 4–12.

Radin, D.I. (2004), 'On the sense of being stared at: An analysis and pilot replication', *Journal of the Society for Psychical Research*, **68**, pp. 246–53.

Rosenthal, R. (1979), 'The "file drawer problem" and tolerance for null results, *Psychological Bulletin*, **86**, pp. 638–41.

Rosenthal, R. (1991), *Meta-Analytic Procedures for Social Research*, Rev. ed. (Newbury Park, CA: Sage).

Scargle, J.D. (2000), 'Publication bias: The "file-drawer problem" in scientific inference', *Journal of Scientific Exploration*, **14** (1), pp. 91–106.

Schlitz, M. and Braud, W. (1997), 'Distant intentionality and healing: Assessing the evidence', *Journal of Alternative Therapies*, **3** (6), pp. 62–73.

Schmidt, S., Schneider, R., Utts, J. & Walach, H. (2004), 'Distant intentionality and the feeling of being stared at: Two meta-analyses', *British Journal of Psychology*, **95**, pp. 235–47.

Sheldrake, R. (1998), 'The sense of being stared at: Experiments in schools', *Journal of the Society of Psychical Research*, **62**, pp. 311–23.

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

Sheldrake, R. (2000a), 'The "sense of being stared at" does not depend on known sensory clues', *Biology Forum*, **93**, pp. 209–24.

Sheldrake, R. (2000b), 'The sense of being stared at: Effects of blindfolding subjects and giving them feedback', *Journal of the Society of Psychical Research*.

Sheldrake, R. (2003), *The Sense of Being Stared At* (New York: Crown Publishers).