British Journal of Psychology (2004), 95, 235–247
© 2004 The British Psychological Society
www.bps.org.uk

Distant intentionality and the feeling of being stared at: Two meta-analyses

Stefan Schmidt[1]*, Rainer Schneider[1], Jessica Utts[2] and Harald Walach[1]

[1]University Hospital Freiburg, Germany
[2]University of California, Davis, USA

Findings in parapsychology suggest an effect of distant intentionality. Two laboratory set-ups explored this topic by measuring the effect of a distant intention on psychophysiological variables. The 'Direct Mental Interaction in Living Systems' experiment investigates the effect of various intentions on the electrodermal activity of a remote subject. The 'Remote Staring' experiment examines whether gazing by an observer covaries with the electrodermal activity of the person being observed. Two meta-analyses were conducted. A small significant effect size ($d = .11$, $p = .001$) was found in 36 studies on 'direct mental interaction', while a best-evidence-synthesis of 7 studies yielded $d = .05$ ($p = .50$). In 15 remote staring studies a mean effect size of $d = 0.13$ ($p = .01$) was obtained. It is concluded that there are hints of an effect, but also a shortage of independent replications and theoretical concepts.

Parapsychological research has received only scant attention from mainstream psychological research. This is partially due to a lack of any conceptual linkage with the normal body of research. We have argued that careful attention to anomalous findings is important for theory building (Walach & Schmidt, 2004). Another reason for disregarding the parapsychological database is the belief that results obtained in parapsychological experiments are not stable. We examined two sets of studies in meta-analyses, in order to make the claims of parapsychology more amenable to a nonspecialized readership and also to find out what direction further research should take.

One experimental paradigm investigates whether there is any relationship between the intentional efforts of one participant (often called the 'agent') and physiological changes in another remote person (often called the 'receiver'). This assumption is referred to as *distant intentionality*.

In these laboratory studies one participant tries, from a distance, to activate or calm the autonomic activity of another participant. The experimental paradigm is known as *Direct Mental Interaction with Living Systems* (DMILS) (Braud, 1994). A variety of

*Correspondence should be addressed to Stefan Schmidt, Institute of Environmental Medicine and Hospital Epidemiology, University Hospital Freiburg, Hugstetter Str. 55, D-79106 Freiburg, Germany (e-mail: sschmidt@ukl.uni-freiburg.de).

different experiments with different designs and target systems have been explored, mainly by Braud and Schlitz (1991) and a set-up employing electrodermal activity (EDA) as a dependent variable proved to be the most successful.

In such a *direct mental interaction* experiment participants come to the laboratory in pairs. One participant is housed in a sound-proof chamber and EDA, as a physiological indicator of general arousal, is recorded continuously. There is no physical stimulation whatsoever. The electrodermal signal is fed back to the second participant, who is placed in front of a computer monitor in another room. The task of the second participant is to either activate or calm the other person by means of intentions only. Several epochs with varying conditions (activate, calm or rest) are presented in a randomized and balanced order. Epochs usually last 30–60 s. A whole session normally consists of 10 activate and 10 calm epochs interspersed with rest intervals. For evaluation purposes, the EDA data of the calm condition are compared with those of the activate condition. A significant difference is interpreted as a so-called *psi* effect, as it reflects a relationship between the physiological condition in one participant and the intentions of the other participant that cannot be explained by conventional means.

A very similar experimental paradigm is known as *remote staring* (Braud, Shafer, & Andrews, 1993a, 1993b). In these studies, the experimental set-up is similar to the experiments on distant interaction described above with one difference. Instead of an EDA feedback, one participant observes the other participant via a closed-circuit television system. Activate and calm epochs are replaced by observation and control epochs depending on whether or not the remote starer observes the other participant on the monitor. Again the EDA data of the observed participant compared for the two conditions. As the observed participant has no knowledge of the observation schedule, no differences would be expected between the two conditions. The *remote staring* paradigm investigates whether humans can detect someone staring at them. This idea is derived from a 'feeling of being stared at' which is well known from everyday experience. This phenomenon has already been researched by individual authors (e.g. Colwell, Schröder, & Sladen, 2000; Schwartz & Russek, 1999; Sheldrake, 1998, 1999, 2000, 2001a), and the results have evoked some controversy (Baker, 2001; Marks & Colwell, 2001; Schmidt, 2001; Sheldrake, 2001b). The *remote staring* paradigm is special in its set-up in that it uses video equipment and takes as the dependent variable a physiological indicator of arousal rather than a self-report.

In an overall review, Schlitz and Braud (1997) report the results of 19 *direct mental interaction* studies with a total of 417 sessions. These show a mean effect size of $r = .25$ (Rosenthal's r) with 37% of the studies being independently significant. The authors also present data for *remote staring* studies with the same mean effect size $(r = .25)$ for 11 studies with 241 sessions, with 64% yielding significant results. The overall significance for the two data sets calculated using the Stouffer Z method was $p = .0000007$ (*direct mental interaction*) and $p = .000054$ (*remote staring*), respectively.

One could conclude that the data indicate the existence of some as yet unknown effect of distant intentionality. Our goal was to assess whether this strong claim can stand up to critical evaluation from a sceptical perspective. We also wished to include new data in our evaluation, as a substantial number of *direct mental interaction* experiments have been conducted since the previous review.

The crucial point in interpreting the results of *direct mental interaction* and *remote staring* experiments is their methodological quality. Only if appropriate safeguards and state-of-the-art methodology are applied to exclude conventional explanations or bias, can one draw conclusions on the phenomenon under research. Earlier analyses of these

experimental studies (Schmidt, Schneider, Binder, Bürkle, & Walach, 2001; Schmidt & Walach, 2000) have shown that techniques in EDA recording, statistical methods and the randomization of the epoch sequence are crucial points for the evaluation, and not all studies have always applied the best techniques available.

Therefore, the main purpose of our meta-analyses is twofold. First, we wished to determine whether there is an overall effect in all *direct mental interaction* or *remote staring* studies. Second, we wished to find out whether effect sizes are related to methodological quality, i.e. whether poor methodological study quality, and thus bias, can account for positive findings. In addition, we formulated a set of exploratory questions regarding moderator variables for the alleged effects.

Methods

Design

Two separate meta-analyses for *direct mental interaction* and *remote staring* studies were conducted. Unless otherwise stated, both analyses applied the same procedures and methods. The scope of the analyses was to cover all the experimental data available, not only published material, as we wished to base our evaluation on the largest available database. To further pool the effect sizes an adaptive procedure was predetermined that allowed the statistical procedures to be changed according to the structure of the data retrieved. The design of this study, the hypotheses and the methods were established beforehand and were conducted according to a predetermined protocol.

Inclusion criteria

We restricted our analysis to experimental studies on distant intentions or 'the feeling of being stared at', respectively, with EDA as the dependent variable. We included all studies that had completed their data collection by the end of 2000. (To our knowledge no new studies have been published between then and September 2003. Thus, this database is still up to date.) We also included both published and unpublished studies.

Literature search

Studies identified in the review by Schlitz and Braud (1997) formed the basis of our data set. A hand search of the seven relevant journals was conducted to include new studies.[1] In searching for unpublished material we contacted all first authors of all the published studies to ask for assistance. Furthermore, we posted a search request on an e-mail forum that discusses parapsychological research issues.

Coding

To code the study characteristics we compiled our own coding form. This was done by extracting all the characteristics mentioned in all the studies included in the analyses. In doing so, we obtained a first item list. Items were checked for usefulness, sorted and grouped. Items were coded directly as exact figures where possible (Stock, 1994). In all other cases, either a list of possible specifications was provided (e.g. brand of

[1] *A list of journals can be obtained from the first author.*

electrolyte applied) or a *Yes – No – Unclear/Unknown* convention was offered. The initial coding form was discussed among SS, HW and a research assistant to avoid ambiguities and to solve difficulties. Next, the form was assessed in a pilot coding. The form was revised and reassesed in pilot coding. After a third pilot coding, we considered the form to be sufficiently optimized for an unambiguous, unbiased and accurate coding procedure. The resulting coding form had 208 items. It can be obtained from the authors on request.

The coding itself was conducted by a single coder (SS) who possesses considerable expertise according to a procedure proposed by Orwin (1994). SS gained his expertise from his own research work, from personal contact with most of the other researchers and from intensive research of the literature (Schmidt, 2002, 2003; Schmidt & Walach, 2000). This kind of coding also included the use of personal information by the coder in cases in which ambiguities in the papers could be solved by SS's personal knowledge. This procedure has the advantage of employing insider knowledge to minimize the amount of missing data and erroneous coding. However, it has the disadvantage of using an unblinded coding procedure that might bias judgment. We chose expertise coding to obtain complete and correct data and tried to minimize any coder bias by a process of continuous self-reflection and control through a second independent coder (RS).

Studies were coded in a randomized order within a short period. At the end of the coding procedure, SS rated each of the 208 items according to their susceptibility to biased or ambiguous coding on a scale of 1 – 5, with 5 being the most susceptible. Five items were rated as 4 or 5. Next, the second coder (RS) independently coded all studies for these five items. The second coder served only as a control for a reduced set of items in order to minimize the work load.[2] The two codings of these most difficult items yielded an agreement rate (Orwin, 1994) of $AR = .80$. Divergences between the two coders were then discussed for consensus; as a result of this one item was dropped completely.

Data extraction

For each experiment, test statistics (t-, z- or p-value) and sample sizes were extracted for each dependent variable. All data were double-checked. Next, for each experiment and variable an effect size (see below) was calculated. If any study had more than one dependent EDA variable a mean effect size for these variables was computed.

Moderator variables

After all studies had been coded, a frequency analysis was performed for all items. Variables showing no or only very little variance were excluded from the analysis. From the remaining variables, important study characteristics (such as date or sample size) were extracted. Next, a set of moderator variables was defined by single items or the combination of several coding items according to the pre-stated exploratory hypothesis. (These were: type of experiment—DMILS or remote staring, date, relation between agent and receiver, laboratory, distance between the two participants, measurement of

[2] *A complete coding of all 208 variables in the 55 identified studies consists of 11,440 single codings. Most of them are on clear-cut facts, such as year of publication, number of experimenters or distance between agent and receiver. With the reduced number of items only 1,275 codings had to be done.*

skin conductance or resistance, measurement of phasic or tonic parameters, length of epochs, epoch conditions, size of study, feedback for agent yes–no).

Finally, we selected all items from the remaining list that referred to qualitative aspects of the studies. These were 19 items. They were grouped into three categories forming three different quality indices:

Safeguard (34 points/100%) with the items 'acoustical and/or electromagnetic shielding of either one (4 points/11.8%) *or* both participants' (8 points/23.5%), 'observing or intentional acting participant is blind to epoch sequence' (4 points/11.8%), 'number of sessions preplanned and completed' (5 points/14.7%), and 'adequate randomization sequence' (17 points /50%).

Quality of EDA methodology (24 points/100%): 'environmental humidity suitable' (2 points/8.3%), 'environmental temperature suitable' (2 points/8.3%), 'type of electrodes suitable' (4 points/16.7%), 'electrolyte suitable' (4 points/16.7%), 'appropriate electrode attachment' (4 points/16.7%), 'sufficient time lag between electrode attachment and start of data recording' (4 points/16.7%) and 'adequate sampling frequency' (4 points/16.7%).

Methodological quality (10 points/100%): 'hypothesis explicitly stated' (2 points/20%), 'number of sessions preplanned' (2 points/20%), 'statistical evaluation preplanned' (2 points/20%), 'study protocol deposited before start of data collection' (2 points/20%), 'suitable evaluation procedure' (2 points/20%).

Most items were rewarded with a maximum of 2 points (*yes* 2 points—*no* 0 points—*unknown/unsure* 1 point). The category *unknown/unsure* was chosen to avoid bias because most studies with nonsignificant results report fewer details about their procedures. Some items were assigned higher weights. These were rewarded with more than 2 points. This was done according to theoretical considerations on the possible impact of the variable. The process of decision was based on a consensus discussion between SS and HW. The highest weight of all items was assigned to the quality of the randomization procedure as this was considered the most important methodological feature. After the weighting all three indices were transformed into a 0–100 scale.

We preferred to assess the relationship between study quality and effect size on the basis of these aggregated indices rather than for single items for the following reasons: (i) this procedure reduces the number of analyses and thus the likelihood of chance findings, (ii) the aggregated scales show more variance than single items, (iii) the weighting procedure reflects the possible impact of variables on the studies' result, and (iv) some sets of single variables are already highly correlated (e.g. some laboratories always applied the same safeguard procedures).

Finally, we wished to determine an index for *overall study quality* as this is a standard procedure for meta-analyses that focus on quality (Stock, 1994; Wortman, 1994). We computed this index by adding the three indices. Thereby, *safeguards* (40%) and *EDA quality* (40%) entered the overall quality scale with double weighting as they were considered to contain the most relevant features, whereas the third index *methodological quality* (20%) provides only additional aspects which in part are already reflected in the ratings of the other indices.

Publication bias

To avoid publication bias, the scope of the present meta-analyses also included unpublished material (Macaskill, Walter, & Irwig, 2001). But this procedure relies on the

fact that a majority of all unpublished studies could be found. As this cannot be assumed, we tested the distribution of the effect sizes for publication bias. The funnel plot obtained (Dickersin & Berlin, 1992; Duval & Tweedie, 2000) was visually inspected and a significance test was performed on publication bias (Begg, 1994).

Effect size

All experimental data stemmed from within-subject designs. Most studies provide *t*-values stemming from either a *t*-test for dependent data or a single mean *t*-test against mean chance expectation. In these instances, we calculated $d = t/(\mathrm{df}^{1/2})$ according to Rosenthal (1994). Other studies display only *z*-scores as they employed nonparametric statistics (e.g. Wilcoxon test, randomization test). Here, the effect size was computed by the isomorphic formula $d = z/(N^{1/2})$ (see Rosenthal, 1994, eqn 16-2). The error variance for each effect size was estimated by $s^2 = 1/N$, which is the variance of d under the null hypothesis, and is a good approximation for the variance of d for small effect size.

Statistical model

It was pre-specified to first try a fixed-effects model on the data. If this was not applicable a random-effects model should be employed. Therefore, we performed a test of homogeneity on the data set (*Q*-statistic, see Laird & Mosteller, 1990) and calculated the amount of variance that could not be explained by the single study's error variances (Shadish & Haddock, 1994, p. 275). All effect sizes were weighted by size of study. Confidence intervals were computed from the standard deviations of mean effect sizes according to Shadish and Haddock (1994). All calculations were performed on Microsoft Excel '97. The integrity of the algorithms was cross-checked by calculations of formulas by hand. Correlations and frequency analyses were computed using SPSS 9.0 for Windows.

Correction

Whereas *direct mental interaction* studies hypothesize and report directed deviations in EDA, *remote staring* studies do not. The hypotheses in the latter are two-sided, requiring only that the data of observation epochs differ from the data of control periods. Thus, the necessary assumption of the random error to be normally distributed with an expected value of zero could not be maintained. We therefore calculated a correction factor by a permutation test that was exactly tailored to this data set. A more detailed description of this procedure can be obtained from the first author.

Results

Direct mental interaction

Forty studies were included in the analysis, 25 of which had already been published when the analysis was performed. The studies were conducted between 1977 and 2000. The characteristics of the data set are shown in Table 1.

The data set just passed the test on homogeneity ($\chi^2 = 53.17$, $\mathrm{df} = 39$, $p = .065$) but not all the variance in the data set could be explained by the error variance in the single studies. Therefore, we decided to conduct a moderator analysis. Correlations

Table 1. Characteristics of the *direct mental interaction* data set

Number of studies	40
Number of sessions	1055
Range of sessions per study	10–74
Range of effect sizes	−0.32 to 1.02
EDA quality	57 (22)
Safeguards	75 (25)
Methodological quality	65 (21)
Overall quality	66 (19)

Note. Quality measures are the means over all studies with standard deviations in parenthesis. Quality measures are scaled on a 0 to 100 scale.

between effect size and six predefined potential moderators were calculated. We found significant negative correlations (Pearson's r, $N = 40$) for *overall quality* ($r = -.43$, $p < .01$), *safeguards* ($r = -.53, p < .01$) and *date* ($r = -.48, p < .01$). Entering these variables in a stepwise regression analysis showed that most of the variance could be explained by the variable *safeguards* ($\beta = -.53, p < .001$), while the other two did not enter the model.

Within the variable *safeguard* the quality of randomization proved to be especially crucial for the correlation ($\beta = -.51, p = .001$). Inspection of the data showed that four experiments did not fulfil the minimum requirements of an acceptable randomization process. These four studies were excluded from the analysis. The remaining data set still showed significant correlations with the moderator variables *overall quality* ($r = -.40$, $p < .05$), *EDA quality* ($r = -.35$, $p < .05$) and *safeguards* ($r = -.37$, $p < .05$). A regression analysis showed that from these variables only *overall quality* was a significant predictor ($\beta = -.47, p = .002$) for effect size.

Hence, higher *overall study quality* was related to lower *effect sizes*, which indicates that some of the reported effects might be due to methodological shortcomings. This was possibly accounted for by either calculation of a mean effect size that was not only weighted by study size, but also by overall study quality or calculation of a best-evidence synthesis (Slavin, 1995), i.e. a meta-analysis of only the studies with a high overall quality index. Both models were applied.

Model 1: Effect sizes weighted by overall quality

The remaining data set of 36 studies (1015 single sessions) passed the test of homogeneity ($\chi^2 = 33.27$ df $= 35, p = .55$) and all variance in the data set could be explained by the error variance in the single studies. It yielded a mean effect size of $d = .11$ ($p = .001$). The 95% confidence interval (CI) ranged from 0.04 to 0.17.

Neither the visual analysis of the funnel plot (Fig. 1) nor the corresponding significance test ($z = .93, p = .18$) showed any indication of a publication bias.

Figure 2 shows the relationship between overall quality and effect size. As a result of the weighting procedure, the correlation between these two variables dropped to a nonsignificant coefficient of $r = -.29$ ($p = .08$).

Model 2: Best-evidence synthesis

The distribution of the overall quality score displayed in Fig. 2 clearly showed a distinct subset of seven high-quality studies in the data set which were chosen for the

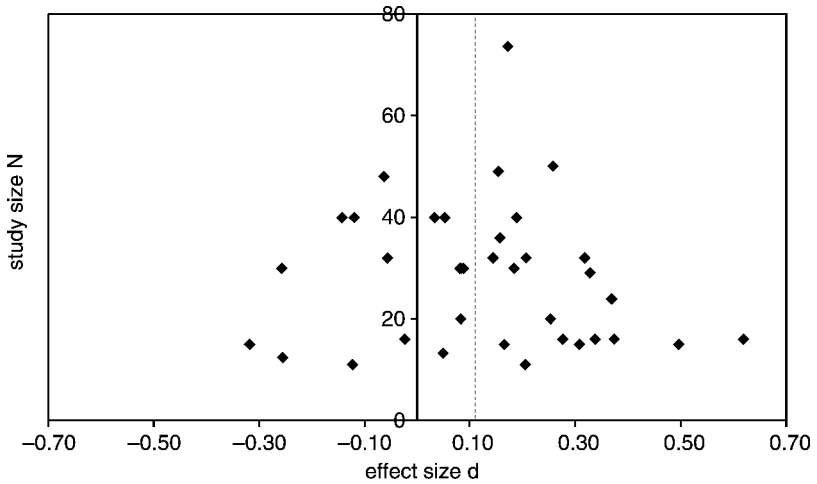

Figure 1. Funnel plot showing the distribution of the effect sizes against study size ($N = 36$). The dotted line indicates the mean effect size $d = .11$.

best-evidence synthesis. These studies have an average overall quality score of 99 and represent with 188 sessions 19% of the overall data. All of them were conducted in the same laboratory by the same research group.

For the best-evidence synthesis studies were only weighted by size. The analysis yielded a nonsignificant mean effect size of $d = .05$ $(p = .50)$. The 95% CI ranged from -0.09 to 0.19.

Remote staring

Fifteen experiments with an overall total of 379 sessions fulfilled the inclusion criteria, 13 of which had been published at the time of the analysis. The studies were conducted between 1989 and 1998. Further characteristics of the data set are displayed in Table 2.

The data set proved to be homogeneous ($\chi^2 = 13.04$, df $= 14, p = .60$). Therefore, no further moderator analyses were conducted. There was a nonsignificant positive

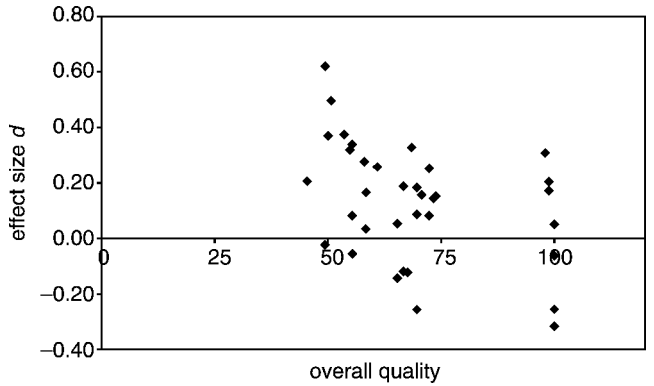

Figure 2. Plot showing the distribution of overall study quality and effect size. Overall quality data was transformed onto a 0 to 100 rank scale. $N = 36$.

Table 2. Characteristics of the *remote staring* data set

Number of experiments	15
Number of sessions	379
Range of sessions per study	15–66
Range of effect sizes	0.04–0.69
EDA quality	49 (17)
Safeguards	67 (10)
Methodological quality	61 (17)
Overall quality	59 (9)

Note. Quality measures are the means over all studies with standard deviations in parenthesis. Quality measures are scaled on a 0 to 100 scale.

correlation between overall study quality and effect size ($r = .26$). The meta-analysis yielded a mean effect size of $d = .28$, which dropped to $d = .13$ ($p = .01$) after correction for confounding with sampling error (see *Methods*). The 95% CI ranged from 0.03 to 0.23.

Discussion

Our meta-analysis examined two closely related experimental paradigms assuming direct interaction between two distant persons.

Meta-analysis is the state-of-the art method for the statistical integration of individual research results. The method implies a high degree of objectivity and reliability. However, any practical application of meta-analysis requires a number of decisions on procedural details at all stages. Nonetheless, it is clear that different procedures can lead to different outcomes (Fishbain, Cutler, Rosomoff, & Rosomoff, 2000; Sacks, Berrier, Reitman, Ancona-Berk, & Chalmers, 1987). Because of the unconventional claim of the studies under research, we always chose a more conservative strategy whenever such a decision had to be made.

Direct mental interaction

The results of the analysis are somewhat ambiguous. Although the quality weighted analysis of 36 studies (model 1) yields a highly significant result, the best-evidence synthesis of seven studies (model 2) does not. In both cases the effect sizes obtained, of $d = .11$ and $d = .05$, respectively, are small according to Cohen (1988). These effect sizes differ considerably from the one ($r = .25$) reported in an earlier review by Schlitz and Braud (1997). Further analyses showed that this reduction is due to: (i) weighting the studies by sample size, (ii) the exclusion of four methodologically weak studies, (iii) weighting the studies for overall quality, and (iv) the inclusion of studies conducted after 1996 which produced smaller effect sizes.

Utilizing a conservative and sceptical approach one could take the result of the best-evidence synthesis to claim no effect of distant intentionality; but there are reasons why such an interpretation should be treated with care. (i) All seven studies were conducted by our research group in Freiburg, Germany. Basing the meta-analysis only on these studies would negate all the results obtained by other researchers. (ii) The construction of the quality index was based on our experience in conducting *direct mental interaction* experiments. Taking this index to exclude all other studies from the analysis would be circular. (iii) Furthermore, the lack of significance in model 2 can have various

reasons. The most obvious would be that there is no such effect. However, it might also be possible that this effect exists, but that the analysis does not attain significance because of a reduction of the data set to only 19% of its original size. If the true effect size is about $d = .1$, more data are needed to reach significance. This line of reasoning is underlined by the fact that we conducted a pilot study, which it was planned should not be included in any meta-analysis (Schmidt *et al.*, 2001). This study yielded a medium effect size of $d = .41$, thus fulfilling the criteria for inclusion in the best-evidence synthesis. The inclusion of this study would result in a mean effect size of $d = .09$. This value is close to the result of model 1.

Model 1 yields a highly significant effect size of $d = .11$ $(p = .001)$. We are confident this result is most likely to reflect the true situation. Are there any alternative explanations for this finding other than an effect of distant intentionality? The significance test on publication bias as well as the funnel plot (Fig. 1) indicate that the distribution of the retrieved studies shows no major bias that could favour a false-positive result of our analysis. The moderator analysis has shown a significant correlation between *overall study quality* and *effect size*. This is a strong indication that some part of the reported effects is due to methodological shortcomings. However, the corrected analysis and the remaining data are still significant. One could argue that independent of high or low weights the alleged artefacts are still in the analysis. The crucial question here is whether methodological shortcomings can be responsible for the significant findings even after the elimination of four studies and after the weighting for quality. The correlation coefficient between *overall study quality* and *effect size* in the resulting analysis is nonsignificant ($r = -.29, p = .08$) and accounts for only 8.5% of the variance. Therefore, we do not think that the effect can be explained away by methodological shortcomings.

Thus, the analysis shows that some parts of the reported effects might be caused by artefacts, but these artefacts cannot completely account for the observed effects. Even though model 2 leaves some room for doubt, model 1, which is in itself conservative, shows that it is difficult to deny an effect of distant intentionality outright. Several known forms of bias have been excluded and cannot be an explanation. Model 2 cannot be the final answer because it uses data from only one laboratory.

We therefore believe that the final conclusion can only be drawn on the basis of several independent high-quality replications of the experiment at different sites by different experimenters. All data in this analysis stems from various laboratories within the parapsychological research community. Unfortunately, so far, other psychophysiological researchers have not examined this paradigm. We hope these data provide enough incentive for them to do so.

It could be shown in this analysis that a high-quality study in this experimental paradigm is characterized by state-of-the-art EDA recording and most importantly by an adequate randomization (see Schlitz *et al.*, 2003 for a detailed description of recommended methodological details including randomization procedures). Proper shielding and preclusion of conventional channels of communication should be a further requirement.

Remote staring

The remote staring meta-analysis shows a non-significant relationship between *overall study quality* and *effect size*; and the set of 15 effect sizes also passes the test of homogeneity. Therefore, the results are more clear-cut. There is a small, but significant, mean effect size of $d = .13$ $(p = .01)$. It can be interpreted as an indication for the

existence of a remote staring effect. Furthermore, one has to consider that both paradigms, *direct mental interaction* and *remote staring*, are very similar. Taking this into account, the result of the remote staring meta-analysis can be regarded as a validation of the results of model 1 in the *direct mental interaction* meta-analysis and vice versa, making the existence of the effect more likely.

However, one has to bear in mind that the *remote staring* data set has a different structure regarding quality scores (see Table 2). The mean *overall quality* of the studies is lower (59%) than in the *direct mental interaction* data (66%) and, most importantly, the study with the highest quality reaches only 71% of the maximum score compared with 100% in *direct mental interaction*. Therefore, taking into account the findings of best-evidence synthesis in the *direct mental interaction* data set, one has to be careful when interpreting the *remote staring* data because there is a lack of high-quality studies and such studies may reduce the overall effect size or even show that the effect does not exist.

Thus, for the *remote staring* paradigm, we conclude that although there is some hint of an anomalous effect in the data, several independent high-quality replications are needed to determine whether these hints represent an unknown phenomenon or are just an artefact.

Conclusion

We conclude that for both data sets that there is a small, but significant effect. This result corresponds to the recent findings of studies on distant healing and the 'feeling of being stared at'. Therefore, the existence of some anomaly related to distant intentions cannot be ruled out. The lack of methodological rigour in the existing database prohibits final conclusions and calls for further research, especially for independent replications on larger data sets. There is no specific theoretical conception we know of that can incorporate this phenomenon into the current body of scientific knowledge. Thus, theoretical research allowing for and describing plausible mechanisms for such effects is necessary.

Acknowledgements

This study was funded by a grant from the Institute of Noetic Sciences, Petaluma, CA. Stefan Schmidt's work was sponsored by the Institut für Grenzgebiete der Psychologie und Psychohygiene e.V., Freiburg, Germany. Stefan Schmidt, Harald Walach and Rainer Schneider receive grants from the Samueli Institute, USA. We would like to thank Marilyn Schlitz for her generous support and Ulli Biedermann for help in developing the coding system.

References[3]

Baker, R. A. (2001). Robert Baker replies to Sheldrake. *Skeptical Inquirer*, *25*, 61.

Begg, C. B. (1994). Publication bias. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

Braud, W. G. (1994). Can our intentions interact directly with the physical world? *European Journal of Parapsychology*, *10*, 78–90.

Braud, W. G., & Schlitz, M. J. (1991). Conscious interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies*, *2*, 1–46.

[3] *A list containing the references of all studies included in the two meta-analyses can be obtained from the first author.*

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

Braud, W. G., Shafer, D., & Andrews, S. (1993b). Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection. *Journal of Parapsychology*, *57*, 373–390.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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.

Dickersin, K., & Berlin, J. A. (1992). Meta-analysis: State-of-the-science. *Epidemiologic Reviews*, *14*, 154–176.

Dickersin, K., & Min, Y. I. (1993). Publication bias: The problem that won't go away. *Annals of the New York Academy of Sciences*, *703*, 135–148.

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

Fishbain, D., Cutler, R. B., Rosomoff, H. L., & Rosomoff, R. S. (2000). What is the quality of the implemented meta-analytic procedures in chronic pain treatment meta-analyses? *The Clinical Journal of Pain*, *16*, 73–85.

Laird, N. M., & Mosteller, F. (1990). Some statistical methods for combining experimental results. *International Journal of Technology Assessment in Health Care*, *6*, 5–30.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, *20*, 641–654.

Marks, D., & Colwell, J. (2001). Fooling and falling into the feeling of being stared at. *Skeptical Inquirer*, *25*, 62–63.

Orwin, R. A. (1994). Evaluating coding decisions. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 139–161). New York: Russell Sage Foundation.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1987). Meta-analyses of randomized controlled trials. *New England Journal of Medicine*, *316*, 450–455.

Schlitz, M. J., & Braud, W. G. (1997). Distant intentionality and healing: Assesing the evidence. *Alternative Therapies in Health and Medicine*, *3*, 62–73.

Schlitz, M. J., Radin, D. I., Malle, B., Schmidt, S., Utts, J. M., & Yount, G. L. (2003). Distant healing intention: Definitions and evolving guidelines for laboratory studies. *Alternative Therapies in Health and Medicine*, *9*, A31–A43.

Schmidt, S. (2001). Empirische Testung der Theorie der morphischen resonanz–Können wir entdecken wenn wir angeblickt werden? Kommentar [Empirical test of the theory of morphic resonance – Dowe feel if somebody is staring at us?]. *Forschende Komplementärmedizin*, *8*, 48–50.

Schmidt, S. (2002). *Außergewöhnliche Kommunikation? Eine kritische Evaluation des parapsychologischen Standardexperimentes zur direkten mentalen Interaktion* [Extraordinary communication? A critcal evaluation of the parapsychological experiment on distant mental interaction]. Oldenburg: Bibliotheks- und Informationssystem der Universität.

Schmidt, S. (2003). Direct mental interaction with living systems (DMILS). In W. B. Jonas, & C. C. Crawford (Eds.), *Healing, intention and energy medicine: Research and clinical implications* (pp. 23–38). Edinburgh: Churchill Livingstone.

Schmidt, S., Schneider, R., Binder, M., Bürkle, D., & Walach, H. (2001). Investigating methodological issues in EDA-DMILS: Results from a pilot study. *Journal of Parapsychology*, *65*, 59–82.

Schmidt, S., & Walach, H. (2000). Electrodermal activity (EDA)—State of the art measurement and techniques for parapsychological purposes. *Journal of Parapsychology*, *64*, 139–163.

Schwartz, G. E., & Russek, L. G. S. (1999). Registration of actual and intended eye gaze: Correlation with spiritual beliefs and experiences. *Journal of Scientific Exploration*, *13*, 213–229.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

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. *Rivista di Biologia/Biology Forum*, *92*, 53–76.

Sheldrake, R. (2000). The 'sense of being stared at' does not depend on known sensory clues. *Rivista di Biologia/Biology Forum*, *93*, 237–252.

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

Sheldrake, R. (2001b). Research on the feeling of being stared at. *Skeptical Inquirer*, *25*, 58–61.

Slavin, R. E. (1995). Best-evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, *48*, 9–18.

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125–138). New York: Rusell Sage Foundation.

Walach, H., & Schmidt, S. (2004). Repairing Plato's life boat with Ockham's razor: The important function of anomalies for mainstream science. *Journal of Consciousness Studies*. Submitted.

Wortman, P. M. (1994). Judging research quality. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97–109). New York: Rusell Sage Foundation.

Received 14 April 2003; revised version received 20 October 2003