

A meta-analysis of forced-choice experiments comparing clairvoyance and precognition

Parapsychologists have long been interested in the question of whether clairvoyance and precognition are really different phenomena (see e.g., Morris, 1982). It has been suggested that apparent precognition might actually consist of a person sampling the present environment via real-time ESP and extrapolating from the information to make an informed prediction about future events (see, e.g., Mundle, 1978). If this hypothesis is correct, then clairvoyance studies would be expected to result in higher effect sizes than precognition studies because of the extra calculational step involved in the latter type of ESP task. Conversely, it has been proposed that what appears to be a clairvoyance task may not involve real-time information acquisition but rather consists of precognition of the feedback of the target's identity later received by the participant (e.g., Carington, 1945). In this case, the effect sizes of the two types of ESP task would be expected to be the same and would be expected to be influenced by the same moderator variables, all other things being equal. There could, of course, be various additional models of clairvoyance and precognition that would have different predictions about the relative strengths of the effect sizes. However, only two papers to date (Tart, 1983; Milton, 1998) have reported comparisons of effect sizes between the two types of studies.

Tart (1983) presented an unusual meta-analysis of 85 forced-choice ESP studies. He discarded all studies that were not statistically significant at an alpha of .05 and also excluded all but the most successful subset of data from any study that broke down the overall result into individual participant data or individual condition or run data. By assessing the bit rate per trial, he found that real-time ESP (clairvoyance or telepathy) studies outperformed the precognitive studies ($p < 5 \times 10^{-4}$). However, the reverse held in Milton's (1998) meta-analysis of 78 free-response ESP studies where the mean effect size (z/\sqrt{N}) for the 6 precognition studies (0.34) was higher than that of both the telepathy studies (0.18) and the clairvoyance studies (0.08).

Although the results from Tart's (1983) and Milton's (1998) papers appear somewhat contradictory, both results have problems of interpretation. Milton's (1998) one-way ANOVA to test the effect size differences between precognition, clairvoyance and telepathy was non-significant. The number of precognitive studies is too small to be representative of the latent population of such studies or to give an accurate estimate of their true effect size, which may be much lower than the observed figure. The confidence intervals around the estimated effect sizes of the three types of ESP overlapped considerably, making it easily possible that the true effect size of the population of precognition studies is below that of the real-time ESP studies rather than above as the point estimates suggest. Tart's (1983) result is also inconclusive, but for different reasons. In this paper the study selection criterion may have introduced a bias. If real-time and precognitive studies had the same effect size but precognitive studies tended to have a larger number of trials than real-time ones, the higher level of statistical power in the precognition studies would have enabled more precognition studies with lower effect sizes to reach the .05 alpha cutoff. This in turn would have meant a decrease in the average effect size for the precognition studies. Tart did not report study sizes and this artefact may be a strong possibility, given that precognition trials can be easier to run than clairvoyance trials and hence easier to

collect in large numbers. Moreover, if more real-time than precognition studies reported breakdowns of data into subsets, Tart's policy of including only the "peak performance" data from each study would have similarly inflated real-time effect sizes relative to precognitive effect sizes.

The current meta-analysis aimed to address more directly the question of whether there is any difference between clairvoyance and precognition by comparing clairvoyance and precognition effect sizes in pairs of forced-choice studies in which both types of ESP were examined. The study pairs were expected to use similar procedures when testing the two types of ESP, with the same experimenters and perhaps even the same participants, making any difference in effect size between the two types of ESP task easier to detect. A secondary aim of the meta-analysis was to see whether or not the optimal conditions for the two types of ESP appeared to differ.

Method

Selection of Studies

It was decided in advance to include only studies published between 1935 and 1997 inclusive in: *European Journal of Parapsychology* (and its forerunner, the University of Utrecht Parapsychology Laboratory's *Research Letter*); *Journal of the American Society for Psychical Research*; *Journal of Parapsychology*; *Journal of the Society for Psychical Research* and *Research in Parapsychology*. Although not decided upon in advance, *Proceedings of the Annual Convention of the Parapsychological Association* were consulted for any additional details for papers reported in *Research in Parapsychology*. The first author scanned all papers that looked as though they might contain relevant studies. This is generally considered to be a more effective method of study retrieval than relying upon search engines or inspecting titles or abstracts only (Glass, McGaw & Smith, 1981).

The criteria for determining which studies were eligible for the meta-analysis were designed to include studies that tested both precognition and clairvoyance and to exclude studies where it was unclear that the appropriate comparison could be made. It was decided in advance to exclude any studies by Levy, Sargent or Soal because questions have been raised about the validity of their work.

Studies were included in the meta-analysis if they met the following criteria (examples of candidate studies that did not meet a specific criterion are given so that those who find their studies missing may understand why):

- Both clairvoyance and precognition were tested in trials reported in the same paper, with the authors clearly making a conceptual distinction between targets selected at the time of the participant's guess and those selected after the participant's guess. Stuart's (1941) Studies P and S were excluded because they were not directly comparable and not originally designed to compare clairvoyance and precognition.
- Papers reported only clairvoyance or precognition trials but stated that the procedure was chosen either to be similar to the procedure used in a previous paper studying the other type of ESP or for the purposes of direct comparison with the other type of ESP examined in the previous paper.

A number of exclusion criteria were also prespecified, but additional criteria (listed below in square parenthesis) were added whenever a new study presented an unanticipated challenge.

- Participants aimed for psi-missing under both clairvoyance and precognition conditions (there were no studies in which there was a high aim in one condition and a low aim in the other).
- [Outcomes were merely reported as an aside with no details about procedure.]
- [Outcomes required correction for a stacking effect and did not provide data with which to make the correction.] This criterion excluded e.g., Brier (1967), Fahler (1957) and Freeman (1969).
- [Possible PK effects on ESP success were examined (e.g., checker effects)] This criterion excluded e.g., Palmer (1966), Weiner & Zingrone (1986).
- Either clairvoyance or precognition trials took place only as the occasional run not intended to serve as comparison trials; or if they apparently had been intended for comparison but either one comprised fewer than 100 trials.
- A paper reported a number of ESP procedures tested in a variety of ways without indicating any particular clairvoyance trials as comparable to any specific precognition trials and the number of trials for either ESP type exceeded the other by more than 80% (a larger ratio being assumed to indicate that the study was not conceptually comparing precognition against clairvoyance sufficiently well to be included). This led to the exclusion of e.g., Ryzl (1962), Tyrell (1936).
- The study's procedure meant that the trials could have been conceived as telepathy ones (for example, if someone else could see the target as the participant made a guess). Beloff and Bate's (1971) studies, for example, were excluded on this basis.
- Clairvoyance and precognition data were not reported separately and thus an effect size for each could not be calculated. This led to the exclusion of Carpenter & Carpenter (1967) and Nielsen and Freeman (1965).
- [Some of the results in a study had been reported in another study already included in the meta-analysis] Thus, for instance, Buzby (1967) and Humphrey (1951) were excluded.

Study Definition

The criteria for defining what constituted a study were designed to group precognition and clairvoyance trials into coherent study pairs. By prespecifying how to group the data into studies, it was hoped that any potential bias that might arise in choosing how to divide the data in any given paper would be avoided. The following criteria for considering a set of trials to be part of a study pair were used:

- Clairvoyance and precognition conditions reported in the same paper were considered a study pair, as were precognition and clairvoyance studies reported in separate published papers as long as it was clear that their authors had intended them to be compared.
- When either precognition or clairvoyance trials were presented as separate experiments in a single paper only because the trials were carried out at different dates without any change in procedure, they were treated as a single study.
- When reports included a number of studies from a number of experimenters, experimenters were held constant in both conditions whenever possible, thus giving separate studies per experimenter. Thus where there were reports of studies

from a number of investigators in a single paper, this paper could represent a number of studies for the meta-analysis.

- If a study manipulated moderator variables that were being coded in the present meta-analysis under both clairvoyance and precognition conditions, each moderator category constituted a separate study pair for the meta-analysis. For example, if participants made verbal responses in some trials and written responses in others, that would constitute a study pair with verbal calls and a study pair with written calls.

Coding

The coding criteria are listed in Appendix 1. Once the first author had coded all the studies, the coding criteria were given to the second author, who independently coded the studies in five papers selected from the database. The authors then discussed ambiguities in the coding criteria, textual ambiguities in the papers and any instances of human error. Coding criteria and the codes assigned to individual studies were changed where appropriate, all such decisions being made by the first author. The first author also reviewed the codes assigned to all other studies in the database, making changes where appropriate to ensure that they had been coded consistently with any amended coding criteria.

Next, the first author gave the third author the revised coding criteria and five other papers from the database. The papers had been photocopied with the results sections blanked out and any reference to the study's outcome excised as far as possible so that the third author would be blind as to the studies' results as well as to the first author's coding of them. The first and third authors then compared their codings, and discussed any criterion definition problems, textual ambiguities and human error until they had mutually agreed how each variable in each study should be coded. The first author then checked the coding of all other papers against any amended coding criteria resulting from her discussions with the third author.

All papers containing studies in the database were then prepared for blind coding in the same manner as before and shared out amongst five independent coders with a background in parapsychology, including the second author¹, who all used the final coding criteria (see Appendix 1). All studies in the database were therefore coded at least twice - once non-blind by the first author and once blind by another coder. About a third of the studies had received a yet additional coding by either the second or third author.

Coding reliability

After the five independent coders had coded the studies in the database, the first author thoroughly checked all disagreements between her coding and that of the five independent coders. As a result, the first author altered five of her previous coding decisions on the basis that they represented human error on her part. All other disagreements remained and the first author's self-corrected coding was used throughout as the data for the meta-analysis. Analyses of the reliability of the coding of the whole database were then performed. For the moderator variables it is only possible to report percentage agreement between coders rather than any correlational measure because the codes are categorical rather than ordinal. The generally acceptable standard for inter-coder percentage agreement is 80% (Hunt, 1997) and on

average, both clairvoyance and precognition studies achieved agreements slightly above this figure, with a range from 55% to 100%. The percentage intercoder agreements for each variable are presented in Table 1.

[Insert Table 1]

Table 2 shows the extent of the agreement in coding between the first author and the blind coders for each procedural safeguard. The mean percentage agreement for the procedural safeguard variables was somewhat lower than for the other moderator variables at 71%, with a range of 52% to 90%. Because the safeguard variables were ordinal, with a three-point scale (safeguard not present, not known if safeguard present, safeguard present), it was possible to calculate post-hoc correlations between coders' judgements. Rosenthal (1991) argues that a correlation is a better indicator of coding reliability than percentage agreement, because a high percentage agreement could merely show that two judges share the same bias while coding (for example, of believing that optional stopping is generally unlikely). When the percentage agreements are considered for both moderator and procedural safeguard variables together, the overall percentage agreement is an acceptable 78%.

[Insert Table 2]

Determining the database sample for analysis

Before the database was assembled, it had originally been intended to restrict moderator variable analyses to a subset of the whole database. This subset would include only those studies in which participants were aiming for high scoring (as opposed to chance or below-chance scoring) in both precognition and clairvoyance trials; otherwise, differences in participants' goals across studies would have rendered any comparison between precognition and clairvoyance effect sizes extremely difficult. This subset was to be called the "high aim sample". This latter sample was then itself to have a subset (a "comparable studies" sample) consisting only of study pairs in which the clairvoyance and precognition trials had been coded by the first author as having at least partially similar procedures (see "Study pair procedural similarity" in Appendix 1). Although the comparable studies sample would clearly be the best sample from which to draw conclusions, it was expected to be considerably smaller than the high aim sample and was only to be examined as a check on the validity of conclusions drawn from the high aim sample. However, the comparable studies sample was larger than expected at 22 study pairs and only 7 study pairs smaller than the high aim sample. It was therefore decided after the coding was completed but prior to undertaking any analyses that the comparable studies sample would be the sample used for the meta-analysis. Unless otherwise stated, all further analyses in this paper are conducted on the comparable studies sample only.

Results

All analyses were preplanned unless otherwise stated.

Cumulated outcome

The summary data for the comparable studies sample are presented in Table 3. The same data for the whole database and for the high aim sample are also presented for

the sake of completeness. In all three databases, both the unweighted and weighted Stouffer z s show a highly statistically significant cumulation for the precognition studies. For the clairvoyance studies, in the comparable studies sample and the high aim sample, both types of Stouffer z are statistically significant, although to a lesser degree than for the precognition studies, presumably due to the larger number of trials involved in the precognition studies. Neither the unweighted nor weighted Stouffer z was significant for clairvoyance studies in the whole sample.

[Insert Table 3]

Comparing precognition and clairvoyance effect sizes

The findings from Table 3 indicate that when forced-choice experiments are conducted under similar conditions, the effect sizes for precognition (0.010) and clairvoyance (0.009) are nearly identical. Neither the whole sample nor the high aim sample indicate a different picture, with the t -tests showing no significant difference between the effect sizes within any of the three databases. There appear to be no distributional peculiarities about the differences in precognitive and clairvoyant effect sizes within individual study pairs in the comparable studies database. The precognition experiments had a higher number of trials overall, although in the comparable studies database there was an equal likelihood of a study pair having clairvoyance trials outnumber the precognition ones ($N = 7$) as there was for a study pair having precognition trials outnumber the clairvoyance ones ($N = 7$). Thus the number of precognition trials is greater in this database simply because when precognition experiments do involve more trials than clairvoyance, the number of precognition trials exceeds the clairvoyance ones by a greater margin. In fact, the difference in number between precognition and clairvoyance trials is mostly due to just two studies (Schmidt, 1969; Stuart 1941).

Table 3
Summary data for the comparable studies sample, high aim sample and whole database

	Comparable studies sample (N = 22 pairs)		High aim sample (N = 29 pairs)		Whole database (N = 31 pairs)	
	Clairvoyance	Precognition	Clairvoyance	Precognition	Clairvoyance	Precognition
Total number of trials	108,995	190,100	154,530	214,160	182,596	271,710
Mean no. trials per study	4,954	8,641	5,329	7,385	5,883	8,765
Median no. trials per study	1,497	1,512	1,760	1,760	1,840	1,840
Unweighted Stouffer z (p)	2.68 (0.004)	4.15 (2x10 ⁻⁵)	1.84 (0.03)	4.21 (1x10 ⁻⁵)	1.04 (0.15)	3.99 (3x10 ⁻⁵)
Quality-Weighted Stouffer z (p)	2.81 (0.002)	4.78 (9x10 ⁻⁷)	2.18 (0.01)	4.73 (1x10 ⁻⁶)	1.58 (0.06)	4.66 (2x10 ⁻⁶)
Mean effect size (SD)	0.009 (0.046)	0.010 (0.050)	0.007 (0.045)	0.013 (0.043)	0.005 (0.040)	0.010 (0.040)
t _{diff} between mean effect sizes	p(2-t) = .65, t = 2.08, 21 df		p(2-t) = .48, t = 0.71, 28 df		p(2-t) = .4, t = 0.85, 30 df	

Larger studies would be expected to give more accurate estimates of effect size, given their lower variance, so N-weighted effect sizes were calculated post-hoc. Again, the two mean effect sizes for clairvoyance and precognition were remarkably similar to each other at 0.030 (clairvoyance) and 0.034 (precognition).

It is clear that this database shows no evidence that either clairvoyance or precognition obtains better results. Before moving on to a discussion of whether the two types of task appear to involve different moderator variables, we will briefly discuss whether study quality appears to be an important issue in this meta-analysis.

Procedural safeguards and effect sizes

Studies were assigned points according to whether they had reported procedural safeguards, with a larger number of points indicating higher quality. Clairvoyance studies could obtain between 0 and 16 points, with the actual range assigned being from 1 to 16 (mean = 10.6, median = 10). Precognition studies could gain a minimum of 0 points and a maximum of 12, and the actual range covered this spread (mean = 7.5, median = 7). Table 4 shows the percentage of studies for which the individual safeguards were known to be present, of uncertain status and known to be absent and the correlation coefficients between the presence of individual procedural safeguards and effect sizes in the studies. Negative correlations indicate a tendency for studies with fewer safeguards to have higher effect sizes. No significant negative correlations were obtained, indicating that if study quality did play a role in the meta-analysis, it was not an obvious one. However, the 22 study pairs offer relatively low statistical power for analyses of this kind so the absence of evidence for a relationship between effect size and the presence of procedural safeguards should not be taken as evidence of absence.

[Insert Table 4]

Moderator variables and effect size

Because it was not known in advance what the characteristics of the database would be and, therefore, which analyses would or would not be possible, the precise statistical analyses on the moderator variables were not preplanned. No analysis was conducted when a comparison group contained five or fewer experiments. The results of the moderator variable analyses are summarised in Table 5. No statistically significant effects were found.

[Insert Table 5]

Interaction analyses

To gain a better idea of whether there was any interaction between type of ESP and any of the moderator variables, post-hoc mixed ANOVAs were performed on variables that had more than five study pairs in each condition. A significant interaction was found between ESP type and the way in which the ESP trials were interspersed ($F(1, 16) = 15.037, p = .001$). That is, performance was better on clairvoyance trials when the precognition and clairvoyance trials constituted two separate experiments, whereas precognition trials performed better when they were interspersed with clairvoyance trials. This result may support Tart's (1983) claim that

participants may have a psychological resistance to precognition. If precognition and clairvoyance trials are interspersed with each other the participant may not have the time to appreciate consciously the implications of trying to predict the future. All other interaction analyses were non-significant. All of the analyses are presented in Table 6.

[Insert Table 6 here]

Optimal conditions

In their meta-analysis of forced-choice precognition studies, Honorton and Ferrari (1989) identified post-hoc what conditions appeared to be optimal for effect size. Too few studies in the current database used all of the conditions that they identified as optimal (selected subjects, individual testing, trial-by-trial feedback) for a preplanned contrast with studies meeting their definition of suboptimal studies (unselected subjects, no feedback, group testing). Nevertheless, it may be useful to identify which conditions the present meta-analysis suggests as being psi-conducive for precognition and clairvoyance studies. Unlike the Honorton and Ferrari meta-analysis, none of the **individual** moderator variable analyses in the present meta-analysis were statistically significant. It is possible, though not certain, that the lack of significant relationships could have been due to low statistical power in the present meta-analysis, which contains only 22

Table 4
Percentage of studies with procedural safeguards present, unknown and absent,
and correlations between presence of procedural safeguards and effect size

	Clairvoyance (N = 22)				Precognition (N = 22)			
	Safeguard (%)			<i>r</i>	Safeguard (%)			<i>r</i>
	yes	unk.	no		yes	unk.	no	
Participant access to target prevented during trial	77	18	5	-0.13	-	-	-	-
Experimenter blind	45	50	5	0.04	-	-	-	-
Participant access to records prevented after trial	73	23	5	-0.19	-	-	-	-
Number of trials in study prespecified	59	27	14	-0.03	59	27	14	0.08
Adequate randomisation	64	5	32	-0.20	73	9	18	-0.29
Checking of target records	36	14	50	0.23	45	14	41	0.41
Checking of calls	27	59	14	0.02	32	55	14	0.13
Checking of hits	41	18	41	0.07	45	23	32	0.16
Random sequence entry point selection made blind to calls	-	-	-	-	45	23	32	-0.01
Average				-0.02				0.08

Table 1
Percentage agreement between first author's and blind coding of paired studies
for each moderator variable

Variable	Comparison	Clairvoyance	Precognition
Study pair procedural similarity	58	-	-
Within vs between subjects design	94	-	-
Participant knowledge of trial type	65	-	-
Experimenter knowledge of trial type	84	-	-
Interspersion of precog. and clair. trials	68	-	-
Participant population type	-	81	74
Response modality	-	84	84
Expectation of success mindset	-	61	84
Target type	-	94	97
Target presentation mode	-	81	81
Timing of experimenter's feedback	-	81	81
Participant feedback of each target	-	55	81
Participant told score	-	97	100
Timing of participants' feedback	-	90	87
Individual vs group testing	-	94	90
Participant contact with experimenter	-	100	100
Predicted direction of results	-	61	61
Target-participant distance	-	77	-
Precognitive timespan	-	-	74
Complexity of target selection method	-	-	81
Average	-	81	84

Table 2**Percentage agreement and correlation between first author's and blind coding of studies for each procedural safeguard variable**

Variable	Clairvoyance		Precognition	
	<i>r</i>	% agreement	<i>r</i>	% agreement
Participant access to target prevented during trial	0.46	90	-	-
Experimenter blind	0.75	84	-	-
Participant access to records prevented after trial	0.23	71	-	-
Number of trials in study prespecified	0.68	74	0.70	77
Adequate randomisation	0.64	74	0.66	74
Checking of target records	0.41	65	0.46	55
Checking of calls	0.47	52	0.45	52
Checking of hits	0.51	68	0.68	77
Random sequence entry point selection made blind to calls	-	-	0.70	77
Average	0.52	72	0.61	69

study pairs. This being the case, it is reasonable to use the contrast in effect sizes within each type of ESP task, rather than statistical significance, as a suggestive guide to potentially interesting variables for future research. Using this approach, for the precognition studies, an effect size of roughly 0.040 was chosen post-hoc as a criterion of indicating a possibly psi-conducive procedure, this effect size being approximately the mean effect size achieved by the highest-scoring contrast groups of precognition studies (as shown in Table 5).

There were three variables with a mean effect size of this magnitude (interspersion with clairvoyance trials, individual presentation of targets, trial-by-trial feedback, target selection within seconds). Notably, the last three optimum procedures for precognition tend to occur together; trial by trial feedback involves individual presentation of the targets and in forced-choice precognition experiments trial by trial feedback will often entail that targets are selected soon after the guess. Therefore, even if further research indicates that the relatively high mean effect sizes in these groups are not artefactual, it will be important to determine whether any effects are due to only one or two of the variables rather than all three.

For the clairvoyance studies, a mean effect size of roughly 0.020 was chosen as the criterion for selecting variables that might be psi-conducive on a similar basis as the criterion selection for the precognitive studies. Four possibly psi-conducive procedures were identified by this means (clairvoyance trials conducted separately from precognition trials, motor response, individual target presentation, trial-by-trial feedback and no physical contact with target material). This group of optimum clairvoyance procedures differs from the precognitive one only by the inclusion of motor response and the separation of the two types of trial (precognitive studies necessarily having no participant contact with the target material at the time of the trial). As with the potentially psi-conducive variables identified for precognition studies, however, these variables are only tentatively suggested as worthy of further study because of the lack of statistical significance of the comparison between mean effect sizes in the contrast groups.

Conclusions

This database provides no evidence to support the idea that clairvoyance works better than precognition. When tested under similar conditions - and for the most part with the same participants - the resultant effect sizes for clairvoyance and precognition appear to be remarkably alike. All but two of the study pairs contained exactly comparable studies. Nevertheless, the results from this meta-analysis are generalizable only in relation to the criteria used for determining the studies to be comparable. The current point scheme (see Appendix 1) used to determine the comparability of studies was implemented in the hope of increasing agreement across coders given the inherently subjective nature of this variable. It is possible that other meta-analyses using a different method for coding study comparability may produce different results. However, in the meantime, the burden of proof rests with those who would argue for a difference between effect sizes under real-time and future ESP.

The finding that precognition and clairvoyance differ in performance according to the way in which the experimental trials are interspersed is interesting, but it must be

treated with caution due to the multiple analyses conducted in this paper. It may nevertheless be worthwhile for future studies to examine whether precognition trials do consistently operate better when they are interspersed amongst clairvoyance trials than when they are conducted as entirely separate experiments. Although the database for this meta-analysis provided no significant evidence for psi-conducive moderator variables, it does illustrate that even those moderator variables that may potentially be psi-conducive (i.e., trial by trial feedback, individual target presentation and, for precognition, a short precognitive time-span) are confounded. Therefore, future work could be aimed at teasing these confounding variables apart. It is possible that the lack of significant findings was due to low statistical power, given that the database had relatively few studies and had only a very small overall effect. The high proportion of variables coded in individual studies as “unknown” also indicates that experimental reports need to specify unambiguously what the procedures and safeguards were as the current status of such reports render meta-analyses vulnerable to considerable noise.

The results from this meta-analysis suggest that theories about parapsychological phenomena and future experimental designs should not rest on assumptions that require a difference in precognition and clairvoyance effect sizes. The fact that the precognition studies in the present database involved, on average, more trials than the clairvoyance studies suggests that this difference in sample size might be true of precognition and clairvoyance studies in general, lending support to the criticism of Tart’s (1983) analysis that we offered in the introduction of this paper. If nothing else, this meta-analysis illustrates just how far parapsychology still has to go in understanding the phenomena it investigates.

Table 5
Moderator variable analyses for clairvoyance and precognition studies

Variable	Clairvoyance				Precognition			
	N	ES	t _{diff}	p	N	ES	t _{diff}	p
Trials interspersed	7	-0.010	-1.74	.10	7	0.040	1.29	.22
Trials separated into experiments	11	0.024			11	0.010		
Special and selected participants	7	0.014	0.30	.77	7	0.017	0.18	.86
Unselected participants	15	0.007			15	0.013		
Motor response	7	0.021	0.80	.44	-	-	-	-
Verbal or written response	15	0.004			-	-		
Participants encouraged to succeed	9	0.010	0.17	.86	7	0.010	-0.54	.59
Participants not encouraged	13	0.007			15	0.020		
Individual target presentation	6	0.018	0.73	.48	6	0.039	1.30	.21
<i>En bloc</i> target presentation	15	0.002			15	0.009		
Participant feedback at end of run	6	0.004	-0.59	.57	7	0.012	-1.13	.28
Participant feedback trial-by-trial	6	0.021			6	0.042		
Experimenter feedback at end of study	-	-	-	-	6	-0.005	-0.55	.59
Experimenter feedback after each run	-	-			8	0.005		
Participant touches target	9	-0.001	-0.84	.41	-	-	-	-
Participant does not touch target	13	0.016			-	-		
Target selected seconds after guess	-	-	-	-	6	0.042	1.69	.11
Target selected longer after guess	-	-			15	0.004		

All p-values are two-tailed.

Table 6
Interaction Effects between ESP type and Moderator Variables

Interaction between ESP type and ...	F(df)	p
Interspersion/Separation of trials	F(1, 16) = 15.04	0.001
Special/Unselected participants	F(1, 20) = 1.18	0.29
Encouragement or not of participants	F(3, 18) = 1.28	0.31
Individual/en bloc target presentation	F(1, 19) = 0.35	0.56
Run-score feedback and trial-by-trial feedback	F(1, 10) = 0.18	0.68

References

General references

BELOFF, J., & BATE, D. (1971). An attempt to replicate the Schmidt findings. *Journal of the Society for Psychical Research*, **46**, 21-30.

BRAUD, W. G., & SCHLITZ, M. (1991). Consciousness and interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies*, **2**, 1-46.

BRIER, B. (1967). A correspondence ESP experiment with high-IQ subjects. *Journal of Parapsychology*, **31**, 143-148.

BUZBY, D. E. (1967). Precognition and a test of sensory perception. *Journal of Parapsychology*, **31**, 135-142.

CARINGTON, W. (1945). *Telepathy: An Outline of its Facts, Theory and Implications*. London: Methuen. P.91

CARPENTER, J. C., & CARPENTER, J. C. (1967). Decline of variability of ESP scoring across a period of effort. *Journal of Parapsychology*, **31**, 179-191.

FAHLER, J. (1957). ESP card tests with and without hypnosis. *Journal of Parapsychology*, **21**, 179-185.

FREEMAN, J. A. (1969). A precognition experiment with science teachers. *Journal of Parapsychology*, **33**, 307-310.

GLASS, G. V., MCGAW, B., & SMITH, M. L. (1981). *Meta-analysis in Social Research*. London: Sage Publications.

HONORTON, C., & FERRARI, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology*, **53**, 281-309.

HUMPHREY, B. M. (1951). Introversion-Extroversion ratings in relation to scores in ESP tests. *Journal of Parapsychology*, **15**, 252-262.

HUNT, M. (1997). *How Science takes Stock: The Story of Meta-analysis*. New York: Russell Sage Foundation.

MILTON, J. (1998). Meta-analysis of free-response ESP studies without altered states of consciousness. *Journal of Parapsychology*. In press.

MORRIS, R.L. (1983). Assessing experimental support for true precognition. *Journal of Parapsychology*, **46**, 321-336.

MUNDLE, C. W. K. (1978). Does the concept of precognition make sense? In Jan Ludwig (Ed.), *Philosophy and Parapsychology*. New York: Prometheus.

NIELSEN, W., & FREEMAN, J. (1965). Consistency of relationship between ESP and emotional variables. *Journal of Parapsychology*, **29**, 75-88.

PALMER, J. (1996). External influence on ESP task performance. *Journal of Parapsychology*, **60**, 193-210.

ROSENTHAL, R. (1991). *Meta-analytic Procedures for Social Research*. London: Sage Publications.

RYZL, M. (1962). Training the psi faculty by hypnosis. *Journal of the Society for Psychical Research*, **41**, 234-251.

TART, C. T. (1983). Information acquisition rates in forced-choice ESP experiments: Precognition does not work as well as present-time ESP. *Journal of the American Society for Psychical Research*, **77**, 293-311.

TYRRELL, G. N. M. (1936). Further research in extra-sensory perception. *Proceedings of the Society for Psychical Research*, **44**, 99-166.

WEINER, D.H. & ZINGRONE, N.L. (1986). The checker effect revisited. *Journal of Parapsychology*, **50**, 85-121.

WOLF, F. M. (1986). *Meta-analysis: Quantitative Methods for Research Synthesis*. London: Sage Publications.

Chronological Bibliography of Studies in Meta-analysis

STUART, C. E. (1941). An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *Journal of Parapsychology*, **5**, 99-137.

RHINE, J. B. (1941). Terminal Salience in ESP performance. *Journal of Parapsychology*, **5**, 183-244.

RHINE, J. B. (1942). Evidence of precognition in the covariation of salience ratios. *Journal of Parapsychology*, **6**, 111-143.

HUMPHREY, B. M., & PRATT, J. G. (1941). A comparison of five ESP test procedures. *Journal of Parapsychology*, **5**, 267-292.

MCMAHAN, E. A., & RHINE, J. B. (1947). A second Zagreb-Durham ESP experiment. *Journal of Parapsychology*, **11**, 244-253.

THOULESS, R. H. (1949). A comparative study of performance in three psi tasks. *Journal of Parapsychology*, **4**, 263-273.

- ANDERSON, M., & GREGORY, E. (1959). A two-year program of tests for clairvoyance and precognition with a class of public school. *Journal of Parapsychology*, **23**, 149-177.
- FREEMAN, J. A. (1962). An experiment in precognition. *Journal of Parapsychology*, **26**, 123-130.
- OSIS, K., & FAHLER, J. (1965). Space and time variables in ESP. *Journal of the American Society for Psychical Research*, **59**, 130-145.
- BUZBY, D. E. (1967). Subject attitude and score variance in ESP tests. *Journal of Parapsychology*, **31**, 43-50.
- BUZBY, D. E. (1968). Precognition and clairvoyance as related to the draw-a-man test. *Journal of Parapsychology*, **32**, 244-247.
- NASH, C. S., & NASH, C. B. (1968). Effect of target selection, field dependence, and body concept on ESP performance. *Journal of Parapsychology*, **32**, 248-257.
- SCHMIDT, H. (1969). Precognition of a quantum process. *Journal of Parapsychology*, **33**, 99-108.
- SCHMIDT, H. (1969). Clairvoyance tests with a machine. *Journal of Parapsychology*, **33**, 300-306.
- HARALDSSON, E. (1978). ESP and the defence mechanism test (DMT). A further validation. *European Journal of Parapsychology*, **2**, 104-114.
- HARALDSSON, E., & JOHNSON, M. U. (1979). ESP and the defence mechanism test (DMT) Icelandic study no. III. A case of experimenter effect? *European Journal of Parapsychology*, **3**, 11-20.
- WINKELMAN, M. (1981). The effect of formal education on extrasensory abilities: The Ozolco study. *Journal of Parapsychology*, **45**, 321-336.
- NASH, C. B. (1982). ESP of present and future targets. *Journal of the Society for Psychical Research*, **51**, 374-377.
- CRANDALL, J. E., & HITE, D. D. (1983). Psi-missing and displacement: Evidence for improperly focused psi? *Journal of the American Society for Psychical Research*, **77**, 209-228.
- TEDDER, W. (1984). Computer-based long distance ESP: An exploratory examination. In Rhea A White & Richard S Broughton (Eds) *Research in Parapsychology 1983*, Metuchen, NJ: Scarecrow Press. 100-101
- JOHNSON, M., & HARALDSSON, E. (1984). The defence mechanism test as a predictor of ESP scores. *Journal of Parapsychology*, **48**, 185-200.

HARALDSSON, E., & JOHNSON, M. (1985). The defense mechanism test (DMT) as a predictor of ESP performance: Icelandic Studies VI and VII. *Research in Parapsychology 1985*. London: Scarecrow Press. 43-44.

MCDONOUGH, B. E., WARREN, C. A., & DON, N. S. (1990). Precognition and real-time ESP performance with an exceptional subject: A replication attempt. *Journal of Parapsychology*, **54**, 141-150.

DON, N. S., MCDONOUGH, B. E., & WARREN, C. A. (1992). Psi testing of a controversial psychic under controlled conditions. *Journal of Parapsychology*, **56**, 87-96.

HARALDSSON, E. (1993). Are religiosity and belief in an afterlife better predictors of ESP performance than belief in psychic phenomena? *Journal of Parapsychology*, **57**, 259-73.

Appendix 1

Coding criteria for the meta-analysis

Moderator variables (described in abbreviated form)

Journal: [EJP/Research Letter]/[JASPR/Proceedings]/[JP]/[JSPR/Proceedings]/[RIP]

Study pair procedural similarity: Are the studies' procedures directly comparable? Different experimenters = 1 point; Different feedback conditions = 2 points; Group testing in one study, individual testing in the other = 3 points; Different response mode (e.g., down-through in one study, selecting responses on computer in the other) = 4 points. [Total of 4 or more points = No]/[2-3 points = Partially]/[0-1 points = Yes]

Within vs between subjects design: [Between subjects design (or mixed)]/[within subjects design (i.e., same subjects or substantial majority (80%+) are the same in both experiments)].

Participant knowledge of trial type: [Subject knows only that clairvoyance and precognition will be interspersed (either in runs or trials) and mixed; they won't know which trials are which at time of guess]/[Subject knows when it is a precognitive trial and when it is a clairvoyance one]/[Subject knows studies only as a precognition study]/[Subject knows studies only as a clairvoyance study]/[Separate studies with different subjects]/[Unspecified, ambiguous or other]

A study was coded as "unspecified" only if an intelligent guess was not possible.

Experimenter knowledge of trial type: Does the experimenter know what condition (clairvoyance or precognition) is operating? [Experimenter knows only that will be interspersed and mixed; experimenter need not be present during trials]/[Experimenter knows when precognition and when clairvoyance if interspersed or counterbalanced across runs]/[Separate experiments at different times or experiments radically different in procedure]/[Other or unknown]

Interspersion of precognition and clairvoyance trials: [Precognition and clairvoyance interspersed in runs or blocks of runs]/[Precognition and clairvoyance interspersed in trials]/[Unknown or other]/[Precognition and clairvoyance separate, not interspersed}

Participant population type: [Special subject selected because of their psychic ability or through outstanding performance in prior tests]/[Special population]/[Selected]/[Experimenter]/[Volunteer (paid or unpaid)]/[Children (under 18)]/[Students (college or university)]/[Animals]/[Mixed/unspecified].

If the participant population consisted of 90% or more of one of the above categories then they were coded as belonging to that category. If less than 90% of the participants belonged to one of the categories, they were coded as being a mixed population. However, if participants consisted of volunteers and another category (e.g., children), they were coded as belonging to the other category.

Response modality: [Open matching (key card face up)]/[Blind matching (key card face down)]/[Unspecified, mixed or precognitive matching, or pack matching (in which a response pack is shuffled to match a target pack)]/[Selecting – include computer studies where participant points to screen]/[Calling, verbal]/[Calling, written]/[Calling, unspecified or mixed]/[Mixed or unknown]

If more than response modality was used, the first method used by the participant was chosen.

Expectation of success mindset: [Yes]/[No]

If the experiment was preceded by a talk on parapsychology or if participants were encouraged to talk about their own psi experiences or if participants were encouraged, the study was coded "yes".

Target type: [Symbol]/[Colour]/[Picture]/[Word]/[Number]/[Other or unknown]/[Mixed]

If the targets fell into more than one category, the aspect of the target that the participant had to identify was coded. For example, if targets consisted of coloured symbols and the participant had to identify the target symbol, only this aspect of the target was coded.

Target presentation mode: [One at time]/[DT & variants]/[Sheet layout, i.e., guessing what was on or would appear on a target sheet or group of targets displayed together]/[Mixed, other or unspecified]

Timing of experimenter's feedback: [At end of whole study or experiment]/[At end of subject's participation]/[At end of run or blocks of runs]/[Trial by trial]/[Mixed]/[Unknown, other or ambiguous]

If the end of the run was also the end of the subject's participation, the feedback timing was coded as being at the end of the run.

Participant feedback of each target: [No]/[Yes]/[Mixed or possibly yes]/[Unknown]

If there was feedback only on hits, the feedback was coded as mixed. Run-score feedback only was coded "no".

Participant told score: [No]/[Yes]/[Mixed]/[Unknown]

If rewards were given, the study was coded "yes". If participants knew what the target was, they were coded as also knowing their score.

Timing of participants' feedback: [Never]/[Delayed (include postal results sent back)]/[On completion of study or experiment or series]/[On completing their participation]/[At end of run or blocks of runs]/[At end of each trial]/[Unknown]/[Mixed]

"Feedback" referred either to feedback of the score or of the target as appropriate. If the timing of the feedback was different for the score and for the target, the most frequently given type of feedback was coded for. For example, if participants were told their score at the end of each run and were then shown the targets at the end of a group of several runs, score feedback was coded for. Conversely, if participants were given trial-by-trial feedback of the target identity but only told their score at the end of each run, then target feedback was coded for.

Individual vs group testing: [Individually (include pairs)]/[In group]/[Mixed]/[Unknown or ambiguous]

If participants were tested one at a time in a group, testing was coded as being individual.

Participant contact with experimenter: [Subjects had direct contact with experimenter]/[Subjects had only indirect contact with experimenter (e.g., mail)]/[Subjects had no contact with experimenter]

"Direct contact with the experimenter" meant some sort of social contact with the experimenter, whether face-to-face or on the telephone and whether during the experiment or beforehand.

Predicted direction of results: [Chance]/[Psi hitting]/[Psi missing]/[Unknown, ambiguous or unspecified]

If the experimenter and the participant were aiming in different directions (e.g., the experimenter expecting psi-missing and the participant aiming high), the experimenter's predicted direction was coded. If one condition was predicted to outperform another (e.g., in a psi-hitting direction) but the less successful condition was not explicitly predicted to be at chance, both conditions were coded as predicting psi hitting.

Target-participant distance: [None (i.e., subject touches target or sealed target pack etc.)]/[Same room/vicinity as subject]/[Same building]/[Same city]/[Same country]/[Abroad]/[Mixed]

If the experiment was an Internet experiment with the target located in a server distant from the participant, the distance between participant and server was coded. If the target was stored in the computer that the participant used, the distance was considered to be "none".

Precognitive timespan: [Seconds or less]/[Minute or more]/[Hour or more]/[Day or more]/[Week or more]/[Month or more]/[Unknown, mixed or ambiguous as to when target selection took place (e.g. if not stated whether target selection was between runs or at the end of the experiment)]

Coders were encouraged to make their best guess. If the targets were selected after all the trials had taken place, the time span was coded as being a week or more.

Complexity of target selection method: Complexity of target selection, precognition studies only. [Hand shuffling, die casting etc.]/[Use of mechanical shuffler etc. with rotations etc. decided by random or prespecified means]/[Entry point (in RNT, deck of cards etc.) decided by hand shuffling; or use of PRNG, true RNG or mechanical shuffler]/[Entry point decided by die casting etc. plus complex calculation]/[Use of stock market figures, weather etc.]

Procedural safeguard variables [Abbreviated]

Clairvoyance only

Participant access to target prevented during trial: Possible access to target whilst guessing. [Subject has access to target material by direct touch]/[Subject may have access, e.g., if target in not explicitly tamper-proof container or not explicitly sealed opaque envelope or target could be guessed at or there was poor supervision or screening is poor or DT and cards not in box]/[Subject has no access to target material because screening was solid (e.g. sealed, opaque envelope) or target was in different room. This includes screen-touch matching, other forms of screening or other adequate precautions]

Experimenter blind: [Receiver's experimenter not blind to target]/[Experimenter possibly not blind, e.g. DT or ambiguous. Includes cases where cards stated to be in box, because here experimenter may have put cards in box]/[Experimenter blind or not present]

Participant access to records prevented after trial: Possible access to/modification of records after guessing. [Participant could access records]/[Participant possibly able to access records (for example, if participant could distract experimenter)]/[Participant not able to access records. Includes use of opaque, sealed envelopes when participants supervised. Includes computer studies, although for special subjects the computer should be explicitly stated to be tamper proof]

Both clairvoyance and precognition

Number of trials in study prespecified: I.e. is optional stopping a possibility? [No. of trials not explicitly prespecified and number of trials does not end in 0 or is not exactly divisible by probability of a hit]/[Mixed (i.e. some studies use even numbers of trials and some do not) or the number of trials was prespecified but not carried out as prespecified or appear to be prespecified because the number of trials was the same across conditions]/[Explicitly prespecified number of trials or classroom testing where number of participants cannot be prespecified, but it is prespecified that those present will take part in a prespecified number of trials each]

Adequate randomisation: [Hand-shuffling, die casting without banking board or with pipped dice]/[Die casting with banking board and non-pipped dice, PRNG or RNG without any check of its randomness; mechanical shuffler with no check of its randomness]/[Actual target order checked for randomness, use of random number table (RNT); empirical cross-check (*not* “control” shuffle); PRNG or RNG or mechanical shuffler with check for the randomness of its running either before and/or after the experiment or on target sequence itself. Extensively checked mechanical shuffler]

Checking of target records: [No duplicate recording of target or unclear but seems unlikely that even if there were a duplicate recording that the two target records will have been cross-checked for either experimenter or participant fraud]/[Duplicate, or automatic independent recording of target]

Checking of calls: [No duplicate recording of participant’s calls]/[Observed recording of participant’s calls or unclear but looks likely that there were duplicate records]/[Duplicate or automatic recording (including photographs) of participant’s guesses]

Checking of hits: [Hits not checked independently by two separate people; unsure but unlikely]/[Recording of hits observed or experimenter does double check; unsure but likely]/[Independent or automatic checking of hits]

Precognition only

Random sequence entry point selection made blind to calls: [Person deciding entry point to randomisation or person recording the random numbers not blind to calls]/[Unknown]/[Person deciding entry point is blind to calls (including computer choice of entry point)]

¹ The first author was responsible for designing and carrying out the meta-analysis; the second and third authors were active as consultants for both statistical and conceptual issues. The quality of this meta-analysis has been much improved by the hard work of the independent coders: Caroline Watt, Dagmar van der Neut, Kathy Dalton and Paul Stevens. To them we are indebted. We are also grateful for the input from two anonymous referees and we would like to thank Simon Sherwood for his assistance with the mixed ANOVA analyses. Finally, this research would not have been possible without the financial support of the first author by the Fundação Bial.