The Journal of Parapsychology, Vol. 67, Fall 2003 (pp.)

ARE ESP AND PK ASPECTS OF A UNITARY PHENOMENON? A PRELIMINARY TEST OF THE RELATIONSHIP BETWEEN ESP AND PK

BY CHRIS A. ROE, RUSSELL DAVEY, AND PAUL STEVENS*

ABSTRACT: This study was designed to explore the relationship between ESP and PK performance by testing for both using a common protocol to control for expectancy effects and experimental artifacts. Forty participants completed a computer-based greyhound racing game. Races occurred in 2 blocks of 12, with one block presented as an ESP task and the other as a PK task. Within each block, half of the races were ESP trials and half PK trials, presented in random order. Overall performance was at chance levels for both ESP and PK trials, for true and disguised trials. There were no significant relationships between performance in the 4 conditions. Although paranormal belief did not predict task success, some other individual differences measures, notably state and trait anxiety, showed some promise. The effect of the mild deception used in this study on participants' performance was considered. Work on this topic is under way to explore more directly the effects of deception.

Gertrude Schmeidler (1988) posed the question: "Is it proper to use psi as a general term for ESP and PK? If it is- if they are alike enough to be classed together- is there any need for the separate terms?" (p. 172). Her question makes explicit an assumption that underlies much of the work in parapsychology, albeit rarely stated, that psi is an intrinsically unitary domain within which ESP and PK are complementary expressions of an inherently undifferentiable and integral set of processes (see also Irwin, 1985, p. 44; Thalbourne, in press).

However, few attempts have been made to test this assumption, and most evidence that bears on the question (some of which we overview below) is circumstantial. Given this ambiguity, theorists have been free to adopt a range of positions vis-à-vis the relation between these two phenomena, depending on one's preferred ontology (see Storm & Thalbourne, 2000). These positions range from assuming ESP and PK to be unitary with neither primary (Schmeidler, 1994b), unitary with PK as the basic phenomenon (attributed by Schmeidler, 1994b, p. 229, to Helmut Schmidt), or unitary with ESP the basic phenomenon (as

We gratefully acknowledge the financial support of the Fundaçao Bial (Grant No. 58/00), which has enabled us to conduct this study. An earlier version of this article was presented in August 2002 at the Parapsychological Association 45th Annual Convention, Paris, France.

captured for example in decision augmentation theory (DAT)[1]: see, e.g., May, Utts, & Spottiswoode, 1995), through to a view associated with William Braud (e.g., Braud, 1985) that sees ESP and PK as complementary phenomena that have quite distinct characteristics and thrive under differing conditions.

This latter position is more in keeping with the view of the general public, who seem to draw a sharp distinction between ESP and PK, regarding the former as much more likely and much more conceivable than the latter (see e.g., Broughton, 1991, p. 35; Schmeidler, 1988, p. 175). For example, Louisa Rhine (1963, p. 88, cited in Irwin, 1999, p. 127) reported that although she had over 10,000 reports of ostensible ESP, only 178 cases of PK were on file.

Similarly, if one takes an operational definition, then one may come to regard the two as clearly distinct, because ESP is measured in terms of a participant's *response* that is then compared with a target, whereas PK is measured in terms of *physical change* in some system in accordance with the participant's intention or aim. However, such operational definitions may tell more about the method of testing than about essential features of the phenomenon, and there is typically still sufficient scope for one to class a phenomenon as an instance of ESP or PK (or both) according to one's preference (see Thalbourne, in press). Ultimately, there is yet no reliable insight into the relationship between ESP and PK.

PRIMA FACIE EVIDENCE THAT BEARS ON A COMPARISON OF ESP AND PK FUNCTION

There is some anecdotal evidence to suggest that ESP and PK abilities are related, as some exceptional individuals who have shown impressive ability in one domain have also performed well in the other, such as Matthew Manning (Schmeidler, 1988) and Eileen Garrett (Healy, 1986). Of course it is possible for individuals to excel in disparate areas (such as academic and athletic achievements) without this coincidence suggesting a common underlying cause. More direct evidence of an association is found in the case of Ingo Swann (Schmeidler, 1973), who is reported to have caused a change in temperature inside a sealed thermos flask at a time when he was attempting to "probe" (locate by ESP) the position of a thermometer inside the thermos. Swann produced a similar effect on a separate occasion

[1] Note, however, that DAT does not attempt to account for instances of ostensible macro-PK.

(Puthoff & Targ, 1974), and Alex Tanous is also reported as producing detectable physical effects at the site of an ESP target on trials in which his call was correct (Osis & McCormick, 1980). These examples seem to suggest that attempts to perceive a target by ESP may involve some physical influence that could constitute a form of PK, although at present they are little more than anecdotal.

Empirical Evidence that bears on a Comparison of ESP and PK Function

Notwithstanding these findings from "gifted" individuals, we concur with Irwin (1985) in arguing that exploring patterns of performance across individuals engaged in PK and ESP tasks may be a more promising means of determining if these phenomena are functionally similar or different. It is disappointing, then, to note that surprisingly little empirical work directly compares performance patterns on ESP and PK tasks; indeed, there seems to have been little interest in performance patterns generally. This makes the task of making reliable judgements concerning points of similarity and difference between ESP and PK an almost impossible one if based only on already-published data. Irwin (1985), for example, has noted that although the empirical grounds for the unitarian view based on an interpretation of performance patterns is rather limited, this may be a consequence of "comparatively meagre quantities of data on performance patterns in ESP, [whereas] similar information for PK is deplorably sparse" (p. xx). Although the situation may have improved somewhat in the case of ESP since Irwin's rather damning conclusion, there has been relatively little change in the case of PK (see Roe, 1996). Where performance patterns have been considered, they may be phenomenon-specific, as Schmeidler (1994b) complained: "It has been frustrating to find that PK and ESP experiments seldom bear directly on the same question and that well-replicated results in one area often have no counterpart in the other" (p. 229). More research is needed that systematically considers the effects of the same variables on ESP and PK performance using methodologically comparable tasks.

To identify the variables that offer most promise of bearing on the question of the ESP–PK relationship, we have taken as our starting point Bem and Honorton's (1994, p. 13) claim:

> The correlation between belief and psi performance is one of the most consistent findings in the parapsychological literature . . . and, within the autoganzfeld series, successful performance of novice (first time) participants was significantly predicted by reported personal psi experiences, involvement with medita-

tion or other mental disciplines, and high scores on the Feeling and Perception factors of the Myers–Briggs Type Inventory.

(But see also Milton & Wiseman's [1999] response, which queries the last two of these.) We consider each of these in turn. Given that ESP is not tested here using the ganzfeld method, it is important to see if similar patterns are to be found from more traditional forced-choice testing methods and to see whether evidence is available to suggest a similar relationship with PK performance.

Belief

Palmer (1971) provided a thorough review of early ESP research that considered the effects of belief on performance, beginning with Schmeidler's original extensive series of seven individual- and 14 group-testing studies (Schmeidler & McConnell, 1958). Although only a small percentage of the studies considered in the review could demonstrate a statistically significant sheep-goat effect, Palmer concluded that the overall pattern of performance was consistent with a real but small effect. Subsequent reviews of ESP research have reached similar conclusions (Lawrence, 1993; Palmer, 1978; Schmeidler, 1994a). For PK research the effects of belief are not nearly so well documented. Gissurarson and Morris (1991) noted that "the sheep-goat classification does not seem to have been adequately tested for PK. The results so far are ambiguous, the reports are sketchy, and the number of subjects participating is low" (p. 123). Some large-scale studies have found a sheep-goat effect (e.g., Gissurarson & Morris, 1991; Morris, Dumughn, Gentles, & Grice, 1993) but others have not (e.g., Troscianko & Blackmore, 1983). Narrower measures of belief including sheep-goat attitude toward PK, perceived success, and PK experience seem to be of most promise (but see also Roe, 1996, for a contrary outcome). Von Lucadou (1987) did report a positive correlation between performance and "confidence" ratings, which he suggested "could be perceived as a kind of sheep-goat variable" (p. 413). This is reminiscent of Schmeidler's "Criterion 1," which asks whether participants think they will do well under the conditions of the experiment (see Schmeidler & McConnell, 1958, chap. 6). However, Rubin and Honorton (1971) surprisingly found that belief in ESP but not PK related positively to success at a PK/I Ching task. Thus there may be some commonality in terms of the effects of belief.

Prior Experience

There have been some attempts to use prior experience as a predictor of ESP performance (see Palmer, 1978, p. 159, for a brief re-

view), but this is clearly a less robust effect than is claimed for the ganzfeld (Honorton, 1997). Palmer (1978) found that in only 2 of 15 studies were researchers able to report a significant positive relationship, although items concerned with prior experience had been included in successful composite belief/experience scales in two further cases. For PK research, the picture seems somewhat more promising. For Gissurarson and Morris (1991), prior PK experience gave the only relationship with PK task performance that was consistently positive across their five studies, yielding a cumulative *z* score of 3.03.

Meditation

There is a long history of writings that suggest that psi phenomena can occur as a by-product of practising meditation or some other mental discipline (see, e.g., Honorton, 1977, 1981, for brief summaries). Some researchers have studied this systematically and found that meditators performed better than nonmeditators. For example, Palmer (1978) described four studies that have considered the effects of meditation techniques on forced-choice ESP performance, of which the outcomes of two were difficult to interpret. Of the remaining two studies, one found a significant difference in favour of meditators, whereas the other was nonsignificant but in the predicted direction. Schmeidler (1994a) reviewed a further six studies, of which four gave significant results in favour of an advantage for meditators. There have been a number of studies that have considered the effects of meditation on PK performance with a random number generator (RNG) as the randomness source. Gissurarson (1997) found that five of the eight studies in his review[2] "gave a significant effect related to meditation and/or prior history of meditation" (p. 95), though it should be noted that, of these, two were conducted on single participants and involved no control condition, and another represented a post hoc finding. Nevertheless, the reported significance levels are sufficiently high to suggest that these studies are not simply optimising on chance.

Feeling-Perceiving on the MBTI

Those participants classified as Feeling–Perceiving (FP) types on the Myers–Briggs Type Indicator (MBTI) have been reported to perform significantly better than non-FP types in ganzfeld trials in three large-scale studies (Honorton, 1997). We are not aware of any previous forced-choice ESP experiments that considered MBTI scores. For PK, however, Schmidt and Schlitz (1989) did find more successful PK per-

[2] One of these (Schmeidler, 1973) used a thermistor as the target.

formance with those who scored highly on Feeling and Perceiving scales. In this study retro-PK was investigated, using mailed prerecorded tapes as targets, and the relationship between this type of effect and real-time PK is unclear. Berger, Schechter, and Honorton (1986) also reported that Feeling types generated higher effect sizes (though not necessarily better "hitting") than Thinking types in real-time tasks involving the PK games *Psi Invaders* and *PsiBall* as well as in silent RNG trials. One of us was unable to replicate these effects in an earlier study (Roe, 1996).

We have seen, then, that a modest case may be made to suggest that both ESP and PK task performance may be optimised by the adoption of Honorton's four-factor model. This study considers whether these relationships hold when both ESP and PK are tested for using a standard protocol. However, we would also like to use this opportunity to consider apparent points of departure between ESP and PK, and for this we focus on participant arousal and prevailing geomagnetic conditions.

Arousal

Useful reviews of the relationship between arousal levels and ESP and PK performance are given by Braud (1981, 1985). There he drew on evidence from a variety of sources to present a strong case for ESP performance being facilitated by reduced autonomic arousal, whereas for PK performance there are some indications that better performance is associated with increased arousal. In this study we have taken measures of state and trait anxiety to be corollaries of arousal level. While we are mindful that anxiety may not always be a corollary of arousal, this approach does have the advantage of allowing us to differentiate between "natural" or usual arousal levels and changes in arousal provoked by the experimental situation.

Palmer (1977) equated anxiety with neuroticism and noted that "when experiments involving group testing are eliminated, the remaining studies reveal a highly consistent picture in support of a negative relationship between ESP and neuroticism" (p. 183). He found that 18 of 24 series were in the predicted direction ($p < .02$) and that all 7 that were independently significant were so. This is reminiscent of Thouless's (1951, cited in Stanford, 1977) reflections on his performance in PK tasks in which he felt that anxiety about success was self-defeating. Indeed, Broughton and Perlstrom (1986, 1992) have found a replicable tendency for scores on measures of state and trait anxiety to correlate negatively with PK score in a competitive computer game context. It is thus unclear whether PK performance is facilitated by arousal or inhibited by it.

Geomagnetic Activity

Past research has often shown a relationship between the state of the geomagnetic field and both success in psi tasks and the frequency of spontaneous psi experiences (for reviews, see Persinger, 1989; Wilkinson & Gauld, 1993. This apparent relationship has been more consistent, possibly due to more research, for ESP protocol experiments, in which a negative correlation has typically been found between the magnitude of geomagnetic field fluctuations and psi success/experience. Persinger and Krippner (1989), for example, found higher scoring for dream-based ESP on days of relatively low geomagnetic activity, whereas Tart (1988) and Makarec and Persinger (1987) found similar relationships for laboratory-based ESP and forced-choice card guessing scores, respectively. However, there is some evidence to suggest that this relationship may be affected by other factors. For example, Spottiswoode (1997) reported a suggestive overall negative correlation but found the effect to be restricted to only a limited window of local sidereal time. Radin, McAlpine, and Cunningham (1994) found the predicted positive correlation between target rank and geomagnetic activity for unselected participants, but they found that the relationship was reversed in a second study that involved creative participants (neither correlation was significant, but they differed significantly from one another: $z = 2.09$, $p = .037$, two-tailed, after Clark-Carter, 1997, pp. 331–332).

PK protocol studies have shown less consistency but are also far fewer. Chauvin and Varjean (1990) found that an applied unchanging magnetic field (broadly equivalent to quiet geomagnetic activity) could be used to significantly increase a directional PK effect on a random mechanical cascade but not to reduce it. Gissurarson (1992) found a significant negative correlation between the directional RNG output and the K index. Nelson and Dunne (1987) reported a nonsignificant positive correlation between the aa index and both RNG output and random mechanical cascade distribution.

AIMS OF THE PRESENT STUDY

We have seen that the evidence to date is mixed with respect to whether ESP and PK phenomena could be regarded as unitary on the basis of common preferred conditions. While there are some suggestions that the two phenomena are disparate (e.g., in terms of relaxation levels of participants), yet other evidence suggests similarity (e.g., in terms of personality types). We also share Schmeidler's (1988) concern that at least some apparent differences between ESP and PK in terms of patterns of performance may simply reflect differences in aspects of the

tasks that are not a consequence of the phenomena themselves. For example, with regard to personality variables, little attempt has been made to guard against scepticism/expectancy effects that may covary with personality dimensions and that in turn could impact on motivation and performance; simply labelling a task as PK may reduce the likelihood of success because participants may be more likely to believe that such a task is "impossible".

Overall, insufficient efforts have been made to ensure that ESP and PK tasks are equivalent in all respects other than the type of phenomenon being tested for. Although not straightforward, it is possible to design protocols that are essentially the same (particularly in terms of the participant's experience of them) but that test for either ESP or PK (or indeed both). Furthermore, because of the potential confounding effects of expectancy, it would also be productive to consider situations in which participants are misinformed of whether a particular trial is of ESP or is of PK. Camstra (1973) did manipulate the briefing given to participants in this way. Some were told that the task involved PK, whereas others were falsely told it was a telepathy task, and as a second variable, some were asked to concentrate and some were told not to. Those who did not know they were involved in a PK task did significantly better than those who did, which may be interpreted as support for a scepticism factor.

This study is primarily an assessment of the validity of a novel method for testing for ESP and PK using a standard protocol, but also consists of an exploratory comparison of patterns of performance for ESP and PK tasks to determine points of similarity and difference.

METHOD

Design

This study incorporated a 2 × 2 repeated measures design looking at the effects of task type (ESP vs. PK) and briefing (informed that the task was ESP vs. that it was PK) on the finishing positions of selected computerised greyhounds in a game format. We also intended to conduct exploratory correlational analyses to determine whether task performance in the four conditions covaried systematically with personality and attitude variables. All analyses were planned to be nonparametric and two-tailed.

Materials and Apparatus

A participant information form (PIF) was constructed that asked about basic biographical and contact details. Of particular interest

here, the PIF incorporated a version of Thalbourne and Delin's (1993) Australian Sheep Goat Scale (ASGS; adapted after Roe, 1998); the Keirsey Temperament Sorter (Keirsey & Bates, 1978)—a variant of the MBTI; and both forms of Spielberger's (1983) State–Trait Anxiety Inventory (STAI). The PIF is a generic form that also includes various other questions (e.g., about hypnagogic/hypnopompic experiences) that were not planned to be a focus of this study. Copies of the PIF are available on request from Chris A. Roe.

A computer program was developed by Paul Stevens that makes use of real-time true random versus pseudorandom data to move six greyhounds from the left to the right of the screen, simulating a race. Screen shots of the program are given in Figures 1 and 2. The number of moves is determined by the output so that over successive iterations some greyhounds move closer to the finish than others. The program monitors progress and notes the order in which the dogs cross the finishing line. The program continues until all six dogs have completed the course. The participants' task is simply (in the ESP condition) to select a dog that they would like to own and that they think will do well in the race, or (in the PK condition) to have their dog identified for them by the computer and for them to "will" it to succeed. In either case, the participants "wins" any prize money awarded based on the dog's finishing position. Prize money is used as a simple weighted score based on finishing position (100 virtual pounds for first, £50 for second, £25 for third, no prize money for the other placings). Each race takes between 9 and 10 s to complete. After a series of 24 races, the participant amasses an amount of overall prize money. The time to com-

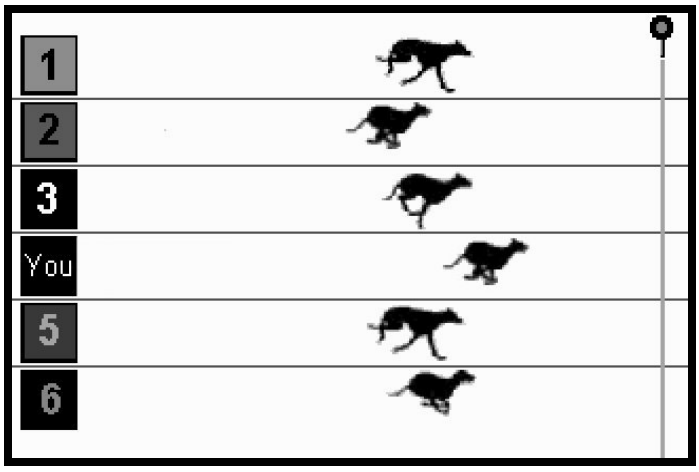

Figure 1. Screenshot of greyhound race

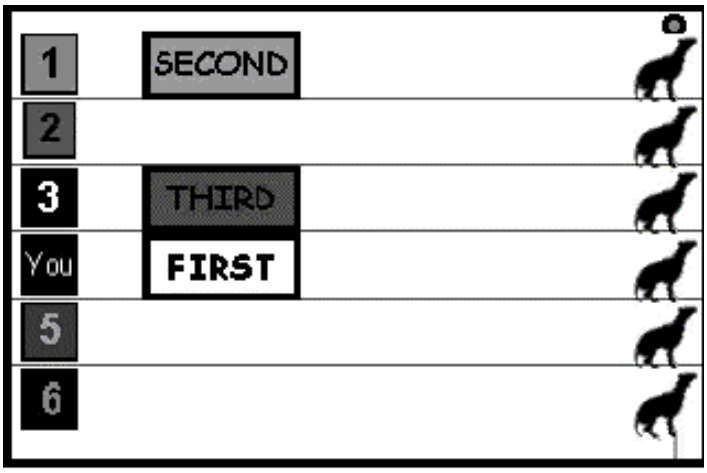

Figure 2. Screenshot of race finish

plete the whole series is dependent on how long participants pause between responses and for this study varied between 9 min 23 s and 14 min 19 s.

A key feature of the project is that all tests make use of a standard task such that the only differences between conditions will be the type of psi required to achieve success and the description of the nature of the task given to participants. Three characteristics distinguish PK trials from ESP trials: source of data, time of data generation, and participant freedom of choice in selecting a target. For PK trials the participants' target dog is selected for them, and races are run in real time with the dogs' movements determined by a highly labile (Orion) electronic noise device; for ESP trials the participants have a free choice of which dog will be their target, the target's movements are determined prior to selection and with reference to a relatively fixed (and theoretically virtually uninfluenceable) set of pseudorandom numbers, and the race is replayed using recorded data. Some trials appear to be ESP but in fact are PK and vice versa. The four conditions were constructed as follows:

True ESP trials: For these the greyhound race was run silently before the trial using pseudorandom data. The outcome was "known" to the program before participants freely selected their greyhound, after which the race was "replayed" on screen

True PK trials: For these the race was run in real time using RNG data. Participants were allocated one of the six dogs using a pseudorandom data file

*Pseudo ESP trials:*Participants apparently "select" one of the six dogs as for the true ESP condition. But in fact the program switches the

data so that whichever they select is exchanged for the one already chosen for them. The trial continues as for the true PK condition

Pseudo PK trials: Again the trial is actually prerun and outcome "known" to the PC. Participants "select" their dog by the timing of their space bar keypress, allowing for a DAT interpretation. Although participants believe they are watching the race in real time, it is in fact a replay.

Participants

Forty people (20 male, 20 female) participated in this study, with a mean age of 27.7 years (*Mdn* = 23.5, *SD* = 10.6). Participants were drawn from an opportunity sample and so consisted mainly of friends and colleagues but also some undergraduate students at the University College Northampton. The sample may be thought of as somewhat sceptical (mean ASGS score = 48.9, *SD* = 14.5, where the theoretical mean score for the scale is 54). Nine had previously participated in a formal parapsychology experiment, and 20 had practised meditation at some time.

Procedure

Prior to the session participants were given the PIF to take away and complete. They were greeted by Russell Davey (RD), who acted as experimenter. In some cases, participants had not completed the measure (e.g., if they had questions about certain items), in which case they were given time prior to their trial to complete the form. Participants next completed the state form of Spielberger's (1983) State–Trait Anxiety Inventory.

They were then escorted by RD into a research cubicle containing a PC with the program ready to begin (six trials were conducted at the homes of participants using a laptop version of the task[3]), and the nature of the task was explained to them and any questions answered. The program autoran and presented participants with a series of 24 races in two blocks of 12. One block was labelled as *gambler* races and consisted ostensibly of ESP trials. Here participants saw the onscreen briefing:

> For the next 12 trials we'd like you to play the role of a gambler who has a free hand to choose which dog to select. In this session the races will already have been run by the computer but not yet have been played out. Your task is to use ESP to identify

[3] There were no significant differences between those trials completed in the participants' homes and those conducted in the laboratory, either in overall performance ($t = .042$, $p = .966$) or for any of the four conditions separately (in all cases, $t < 1.0$, $p > .35$).

> which of the 6 dogs won the race. Once you've made your choice you'll see a replay of the race on screen.

Prior to each gambler race, participants were prompted to enter a number from 1 to 6 corresponding to their choice of dog for the forthcoming replay. A second block was labelled as *owner* races and consisted of ostensible PK trials. Here the onscreen briefing was as follows:

> For the next trials you will play the role of an owner whose greyhounds are entered in a series of races. Your dog will be pointed out at the beginning of each race, and its speed will be determined by a random number generator in the computer. Your task is to try to use PK to influence the RNG so that your preselected dog wins the race. You'll see the race in real time so you get feedback on how well you're doing.

Prior to each owner race, participants were asked to press the space bar to start the race. All of the participants completed both blocks with the order of completion counterbalanced across participants. Within each block, half of the trials were as given in the briefing (e.g., tested for ESP in the gambler block), but half were not (e.g., tested for PK in the gambler block) to gauge the effect of expectation on performance. The experimenter (RD) remained outside the research cubicle during trials but was available should assistance be required. After the program had finished, RD debriefed participants, describing the nature of the four conditions within the task and explaining the need to disguise certain aspects of it. Given the mild deception involved, great pains were taken to ensure that participants were satisfied of the need for the study to be designed as it was and to be sure that they were happy for their data to be included in analysis. No participants asked to withdraw.

Results and Discussion

All analyses presented here are conservatively set as two-tailed unless otherwise specified. Although the primary measure here is the finishing position of participants' greyhounds in computer races, we can get a sense of whether overall performance was above mean chance expectation (MCE) by firstly considering the overall amount won by each participant. The greater the success at the task, the greater the amount of prize money that will have been won. If chance alone is operating, then a participant will typically have won four times in the 24 trials (1/6 likelihood) and have been second and third four times, respectively. This would give total prize money of £700. We can see from Figure 3 that in fact in this study the average prize money is nonsignifi-

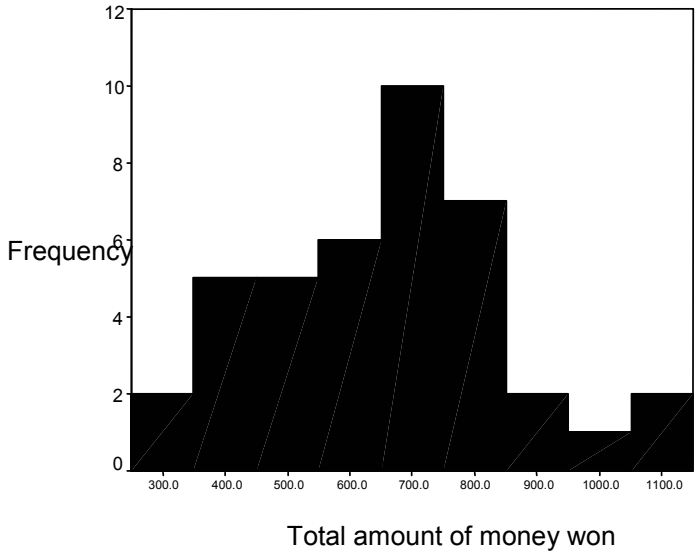

Figure 3. Frequency histogram of prize money "won" by participants

cantly below this, M = £648.10, SD = £199.39; Wilcoxon z = –1.53, N = 40 (3 ties), p = .125.

Given the idiosyncrasies of the distribution of possible scores that arise from the prize-money scale, we planned in advance to use sum of ranks for final finishing position as the principal outcome measure. The distribution of ranks for each of the four conditions is given in Table 1. We can see that in terms of overall scoring, results in this study are generally disappointing. The overall sum of ranks for target dogs is above the MCE of 840 in all four conditions, suggesting that partici-

TABLE 1
SUM OF RANKS (SOR) FOR GREYHOUND FINISHING POSITION

Condition	Finishing position						SOR	z score	Effect size (r)
	1	2	3	4	5	6			
MCE	40	40	40	40	40	40	840		
True ESP	39	28	44	38	49	42	876	–1.34	–.086
Disguised ESP	37	43	40	39	41	40	844	–0.13	–.008
True PK	35	32	46	41	42	44	875	–1.30	–.084
Disguised PK	32	45	38	39	37	49	871	–1.15	–.074
Total	143	148	168	157	169	175	3,466		

pants are faring somewhat worse than chance expectation. None of these deviations is significant, and the effect sizes are small (all Cohen's [1988] *r*s are less than 0.1). There is no difference in performance across the conditions taken together, Friedman's $\chi^2(3, N = 40) = 577$, $p = .90$, and no difference between ESP and PK trials when true and disguised trials are combined, Wilcoxon $z = -.475$, $N = 40$ (1 tie), $p = .64$. Similarly, there is no difference between scores for those tasks perceived as ESP and those perceived as PK, Wilcoxon $z = -.435$, $N = 40$ (2 ties), $p = .663$, failing to support the notion of a "scepticism factor" inrelation to PK tasks. Indeed, the worst performance here was with true ESP. We have therefore not been able to replicate Camstra's (1973) finding that participants in a PK study who were falsely told that they were completing an ESP task fare better than those who were accurately briefed.

The lower than MCE sum of ranks is not simply a result of fewer first placings, but rather it can be seen that there is a tendency for frequencies to increase as one moves from first place through to sixth. Correlating frequency against finishing position gives a significant Spearman's correlation, $r_s(4) = .943$, $p = .005$, suggesting a general shift towards lower ranks.

Notwithstanding this failure to capture an overall above-chance level of scoring, it is still interesting to consider whether similar patterns of performance across individuals are evident for ESP and PK conditions (either informed or disguised). One means by which we may explore this is to see whether individuals' scores are consistent with one another or covary similarly with attitude and personality measures. Correlations of individual sum of ranks scores are given in Table 2. We can see from this that none of the correlations come close to statistical significance, indicating that performance in one condition cannot be predicted on the basis of performance in any of the other conditions. Indeed, surprisingly, the strongest relationship occurs between true ESP and true PK conditions. However, it should be noted that these correlations are not significantly different from one another: for dis-

TABLE 2
SPEARMAN RHO CORRELATION COEFFICIENTS (AND ASSOCIATED PROBABILITIES IN PARENTHESES) FOR COMPARISONS OF INDIVIDUAL PERFORMANCES IN THE FOUR CONDITIONS

Condition	True ESP trials	Disguised ESP trials	True PK trials
Disguised PK trials	.222(.168)	–.038(.817)	–.121(.458)
True ESP trials		.124(.445)	.241(.134)
Disguised ESP trials			.074(.648)

guised PK/true PK versus disguised PK/true ESP, $t(37) = 1.766$, $.10 > p > .05$; for disguised PK/true PK versus true PK/true ESP, $t(37) = 1.85$, $.10 > p > .05$ (after Clark-Carter, 1997, pp. 526–527).

Table 3 gives the correlation coefficients for the relationship between individual differences measures and performance in the four conditions. It is important to note that the outcome measure here is sum of ranks so that greater scores indicate worse performance at the task. Thus positive correlations with belief indicate that higher scores on the belief and attitude measures are associated with worse performance at the task whereas negative correlations indicate better performance at the task as belief scores increase. With this in mind, we can see that the strongest relationship in Table 3 is with prior experience, for which greater numbers of experiences are associated with better performance here, but only the overt PK task—indeed, there is a suggestive trend in the opposite direction where the PK task is hidden. The

TABLE 3
SPEARMAN CORRELATIONS BETWEEN TASK PERFORMANCE AND BELIEF AND PERSONALITY VARIABLES

Condition	True ESP	Disguised ESP	True PK	Disguised PK
PK Criterion 1	.151	.011	.230	.227
	(.351)	(.944)	(.153)	(.159)
Overall ASGS score	.055	-.228	-.103	.054
	(.734)	(.158)	(.527)	(.740)
ESP factor	.115	-.256	-.071	.035
	(.478)	(.112)	(.655)	(.829)
PK factor	-.016	-.180	-.153	.067
	(.923)	(.265)	(.346)	(.681)
Survival factor	-.141	-.004	-.082	.078
	(.384)	(.981)	(.617)	(.632)
Prior experience	-.029	.051	-.433	.269
	(.858)	(.754)	(.005)	(.094)
State anxiety on STAI	.281	-.138	-.185	.317
	(.080)	(.395)	(.254)	(.046)
Trait anxiety on STAI	.272	-.120	.077	.275
	(.089)	(.461)	(.638)	(.086)
3 hr K Index value	.288	-.124	.147	-.263
	(.071)	(.447)	(.364)	(.101)

Note. Probabilities in parentheses are two-tailed. ASGS = Australian Sheet Goat Scale [Note: ASGS has been defined in the materials section] ; STAI = State–Trait Anxiety Inventory.

difference between these correlations is significant, $t(37) = 3.25$, $p < .01$, suggesting that briefing may affect the nature of this relationship. This confirms the findings across five studies of Gissurarson and Morris (1991) but suggests it may be related to expectancy in some way. There is no relationship between prior experience and ESP task scores, which is consistent with Palmer's (1978) review. In attempting to better understand this nonrelationship, we are grateful to an anonymous reviewer for commenting that in the reviewer's experience when debriefing participants, those who reported past ESP experiences tended to associate these with emotional involvement to another person. Forced-choice tasks are typically rather neutral (although we did endeavour to make ours involving) and do not involve another person, and so may lack the ecological validity to convince participants that ESP could occur. (Of course, where the task does seem ecologically valid, then we might expect prior experience to predict success [see Bem & Honorton, 1994, but see also Milton & Wiseman, 1999].) In future studies we intend to more directly measure participants' expectations of success at this task; it will be interesting to see how this relates (if at all) to prior experience.

In terms of belief, our Criterion 1 variable ("I will be able to demonstrate any PK ability that I have in a controlled laboratory experiment") gives stronger correlations with PK performance, but note that these are in the "wrong" direction, with greater confidence predicting *worse* performance.[4] This is in contrast to von Lucadou's (1987) reported positive correlation but could possibly be as much an indicator of performance anxiety as of belief.

Considering the subscales of the ASGS, there is no clear pattern that gives confidence either for or against a view of ESP and PK as aspects of a unitary phenomenon. ESP subscale scores seem to be slightly better predictors of task performance than either of the other subscales, but this is primarily due to a modest correlation with disguised ESP, which in itself is not significant. Clearly, in this study prior belief is not significantly related to performance in any of the psi conditions. This is in contrast to the small but relatively consistent effect of belief on ESP performance described by Lawrence (1993) and others but for PK adds to the rather murky picture, with some authors reporting a sheep-goat effect (e.g., Morris et al., 1993) while others have not (see Gissurarson, 1990/1991). It is interesting to note that there are quite

[4] As with a number of associations reported in this article, these correlations are modest and do not come close to statistical significance, so we must be wary of overinterpreting them here. Given that this is an exploratory first study, we felt it was important to speculate on these findings here on the understanding that the weak effects noted must be confirmed in future planned replications to warrant our continued interest.

strong relationships with both state and trait anxiety and performance at both ostensible ESP tasks, with the effect strongest where the task was actually a disguised ESP one. This broadly replicates the effect reported by Broughton and Perlstrom (1986, 1992), with those reporting greater anxiety performing worse at the task. Again, however, this does not represent a point of difference in the action of ESP and PK.

With regard to geomagnetic activity, there is a suggestive tendency for performance at the true ESP task to be better when activity is low, which is consistent with previous ESP research reviewed by Persinger (1989). However, this relationship does not hold for the disguised ESP trials, for which in fact there is a modest correlation in the opposite direction to prediction. For PK performance the strongest effect is for disguised PK trials and is in the same direction as reported by Nelson and Dunne (1987), although again the direction of relationship is reversed when we consider true PK trials, which show a modest positive correlation.

Finally, we attempted to replicate the tendency for those who present as Feeling–Perceiving (FP) on MBTI measures to outperform those who present as other types. The mean sums of ranks for FPs and non-FPs are given in Table 4. Again, note that higher sums of ranks indicate worse performance at the task. It is clear that there are only modest differences between types on all four tasks and these are swamped by larger within-group differences. None of these approach significance (Wilcoxon *z* scores are in the range 0.142 to 1.071), thus failing to confirm previous findings suggesting superior performance for FP types in ESP tasks (e.g., Honorton et al., 1990) and PK tasks (e.g., Schmidt & Schlitz, 1989; but see also Roe, 1996). There are still too few studies that have considered personality correlates for us to

TABLE 4
MEAN SUM OF RANKS (AND STANDARD DEVIATIONS) FOR FP AND NON-FP TYPES FOR THE FOUR CONDITIONS

Condition	True ESP	Disguised ESP	True PK	Disguised PK	Overall
FP	22.24 (4.31)	21.21 (5.13)	22.43 (3.78)	22.86 (3.78)	88.64 (8.99)
Other	21.77 (3.82)	21.04 (4.42)	21.58 (4.48)	21.19 (3.81)	85.58 (9.45)
Wilcoxin *z*	-.142	-.242	-.726	-1.071	-.852
p (2-tailed)	.900	.812	.474	.292	.408

Note. FP = Feeling–Perceiving.

reach any firm conclusions, but it is fair to say that findings from such research to date do not offer great cause for optimism that they will identify important mediators of PK. With respect to the use of the MBTI in particular, although this has been a preferred instrument for measuring personality attributes in parapsychological research (e.g., Honorton, 1997) and has been widely used in psychological assessment (particularly personnel selection; see Kline, 1993), it has been the subject of sustained criticism (DeVito, 1985; Wiggins, 1989). Psychometricians have objected to its use of forced dichotomies, which contrasts with most psychologists' conception of personality as made up of dimensions that represent continua, with individuals being placed on each continuum according to quantitative differences in that dimension (Anastasi & Urbina, 1997).[5] Empirical evaluations of the measure have also questioned its validity, and it is not readily amenable to factorial explication (see Kline, 1993, for a review). Thus the MBTI seems of little utility as a personality assessment, and we would recommend that parapsychologists adopt other personality measures.

General Discussion and Conclusion

In this study there is little evidence to suggest that ESP and PK performance are related to one another; but then neither were true ESP and disguised ESP nor true and disguised PK very highly correlated. Underpinning any attempt to better understand the correlates and preferred conditions for the action of different forms of psi is the assumption that performance is reasonably consistent. At present, we cannot claim that psi is consistent across time, or even in this case consistent in a split-half type of reliability design. Palmer (1977, p. 176) estimated the average reliability of ESP scores to be around .30, which although better than the .124 correlation reported here still falls some way short of Kline's (1986) minimum requirement of .7 for the reliability of psychometric instruments to be considered satisfactory.[6] For PK, the situation is even worse as we report a negative correlation here, which is reminiscent of Boller and Bösch's (2000) test–retest figures that ranged from .269 to –.045. We share their concern that PK is not sufficiently reliable for us to identify meaningful correlates, and this is one of the major obstacles to furthering our understanding of psi.

One suggestion to account for inconsistency here is that we have no psi and that any apparent pattern is simply due to chance and hence

---

[5] We should note, however, that although not usual, MBTI responses can be scored continuously (see, e.g., Palmer, 1997).

[6] Even this figure may be optimistic. One anonymous referee estimated that reliability estimates for their participants' data were highly variable but tended to hover around 0.10.

should not be expected to vary lawfully. Contributing to this failure to capture psi may be our strategy of recruiting participants who lay no particular claim to strong ESP or PK ability. However, to study covariation, it is essential that we have variation in the first place (to study correlates of IQ, it would not be sensible to only draw participants with IQs above, say, 130), and in this study those variables that might have been used to help screen participants, such as prior belief and experience, have not proved to be especially predictive of success. It may be more useful to point to some suggestive patterns in our data, including some small-to-medium effect sizes that may reflect faculties (whether related or not) that are naturally rather weak and inconsistent and that may have not achieved significance here because of the relatively low power of this study. Because three further studies are planned whose data will be ultimately combined, we hope to be able to determine whether this is the case.

Perhaps participants in this study were not sufficiently motivated. Informal feedback does suggest that despite our best efforts the task can begin to seem repetitive and boring. Perhaps the introduction of an incentive, such as prizes for the highest scorers, could be useful here. But we are mindful of Thouless's (1951, cited in Stanford, 1977) suspicion that increasing motivation may also increase anxiety. We had hoped to alleviate any such anxiety by incorporating a gamelike atmosphere, which Thouless regarded as more likely to be successful. However, this may depends on one's attitude to games, and Broughton and Perlstrom's (1986, 1992) findings suggest that, at least for some, a game format may simply add to the competitiveness of the situation. We thought it would be helpful to provide participants with a goal by letting them know what kind of score was typical so that they might aim to beat it. In retrospect this may have simply contributed to any pressure to perform they might have felt. In future, it may be more useful to emphasise that the task is noncompetitive and pay even more attention to the fact that data are anonymised.

Alternatively, this study might qualify in Storm and Thalbourne's (2000) terminology as "difficult and complex" and thus comprise of more than one inhibiting feature that could hinder success at the task. To some degree this seems to depend on the participants' perception of the nature of the task. Here, we were at pains to present the task as straightforward and intuitive (concerned with winning races rather than on any psi process that might be required). One might argue that participants would have found the PK task more complex, but there is no evidence here that performance was worse than for the ESP task.

Finally, it may be that the lack of overall scoring was a result of participants being aware at some level of the mild deception that was involved in some conditions of this study. Rather than leading to de-

pressed scoring on only those conditions, there may have been a general disenchantment effect (though if so, this occurred at a level that was unconscious and was not reflected in participants' comments during debriefing). Nevertheless, it is an important consideration in parapsychological experiments as to whether one can actually misinform or only partially inform participants in an effort to guard against expectancy effects. It is conceivable that psi is sufficiently boundless to allow participants to be aware of the experimenter's intentions and to react to these rather than to what they have been told. In the second study of this project, we are exploring this possibility by having some participants informed of the true nature of all trials while others experience conditions similar to this experiment. It will be interesting to see whether these groups differ.

References

ANASTASI, A., & URBINA, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

BEM, D. J., & HONORTON, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, **115**, 4–18.

BERGER, R. E., SCHECHTER, E. I., & HONORTON, C. (1986). A preliminary review of performance across three computer psi games [Abstract]. In D. H. Weiner & D. I. Radin (Eds.), *Research in parapsychology 1985,* 1–3. Metuchen, NJ: Scarecrow Press.

BOLLER, E., & BÖSCH, H. (2000). Reliability and correlations of PK performance in a multivariate experiment. *Proceedings of Presented Papers: The Parapsychological Association 43rd Annual Convention*, 380–382.

BRAUD, W. G. (1981). Psi performance and autonomic nervous system activity. *Journal of the American Society for Psychical Research*, **75**, 1–35.

BRAUD, W. G. (1985). ESP, PK and sympathetic nervous system activity. *Parapsychology Review*, **16**, 6–16.

BROUGHTON, R. S. (1991). *Parapsychology: The controversial science*. New York: Ballantine Books.

BROUGHTON, R. S., & PERLSTROM, J. R. (1986). PK experiments with a competitive computer game. *Journal of Parapsychology*, **50**, 193–211.

BROUGHTON, R. S., & PERLSTROM, J. R. (1992). PK in a competitive computer game. *Journal of Parapsychology*, **56**, 291–306.

CAMSTRA, B. (1973). PK conditioning. In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.), *Research in parapsychology 1972,* 25–27. Metuchen, NJ: Scarecrow Press.

CHAUVIN, R., & VARJEAN, B. (1990). Is it possible to strengthen the psi effect using a very weak magnetic field? *Journal of the Society for Psychical Research*, **56**, 96–97.

CLARK-CARTER, D. (1997). *Doing quantitative psychological research: From design to report*. Hove, England: Psychology Press.

COHEN, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

DEVITO, A. J. (1985). Review of the Myers–Briggs Type Indicator. In J. V. Mitchell Jr. (Ed.), *Ninth mental measurements yearbook,* 1030–1032. Lincoln: University of Nebraska Press.

GISSURARSON, L. R. (1990/1991). Some PK attitudes as determinants of PK performance. *European Journal of Parapsychology*, **8**, 112–122.

GISSURARSON, L. R. (1992). The psychokinesis effect: Geomagnetic influence, age and sex differences. *Journal of Scientific Exploration*, **6**, 157–165.

GISSURARSON, L. R. (1997). Methods of enhancing PK task performance. In S. Krippner (Ed.), *Advances in parapsychological research 8,* pp. 88–125. Jefferson, NC: McFarland.

GISSURARSON, L. R., & MORRIS, R. L. (1991). Examination of six questionnaires as predictors of psychokinesis performance. *Journal of Parapsychology*, **55**, 119–146.

HEALY, J. (1986). Hippocampal kindling, theta resonance and psi. *Journal of the Society for Psychical Research*, **53**, 486–500.

HONORTON, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology,* 435–472. New York: Van Nostrand Reinhold.

HONORTON, C. (1981). Psi, attention states and the yoga sutras. In B. Shapin & L. Coly (Eds.), *Concepts and theories of parapsychology,* 55–68. New York: Parapsychology Foundation.

HONORTON, C. (1997). The ganzfeld novice: Four predictors of initial ESP performance. *Journal of Parapsychology*, **61**, 143–158.

HONORTON, C., BERGER, R., VARVOGLIS, M., QUANT, M., DERR, P., SCHECHTER, E., & FERRARI, D. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology*, **54**, 99-139.

IRWIN, H. J. (1985). Is psi a unitary domain? Analysis in terms of performance patterns. *Parapsychological Journal of South Africa*, **6**, 34–46.

IRWIN, H. J. (1999). *An introduction to parapsychology* (3rd ed.). Jefferson, NC: McFarland.

KEIRSEY, D., & BATES, M. (1978). *Please understand me: Character and temperament types*. Del Mar, CA: Prometheus Nemesis Books.

KLINE, P. (1986). *Handbook of test construction*. London: Methuen.

KLINE, P. (1993). *The handbook of psychological testing*. London: Routledge.

LAWRENCE, T. R. (1993). Gathering in the sheep and goats: A meta-analysis of forced-choice sheep-goat studies, 1947–1993. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 75–86.

LUCADOU, W. V. (1987). A multivariate PK experiment: Part III. Is PK a real force? The results and their interpretation. *European Journal of Parapsychology*, **6**, 369–428.

MAKAREC, K., & PERSINGER, M. A. (1987). Geophysical variables and behaviour: XLIII. Negative correlation between accuracy of card guessing and geomagnetic activity. *Perceptual and Motor Skills*, **65**, 105–106.

MAY, E. C., UTTS, J. M., & SPOTTISWOODE, S. J. P. (1995). Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration*, **9**, 453–488.

MILTON, J., & WISEMAN, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, **125**, 387–391.

MORRIS, R. L., DUMUGHN, I., GENTLES, P., & GRICE, L. (1993). *Sheep-goat effects and RNG PK game performance*. Paper presented at the 17th International Conference of the Society for Psychical Research, Glasgow, U.K.

NELSON, R. D., & DUNNE, B. J. (1987). Attempted correlation of engineering anomalies with global geomagnetic activity [Abstract]. In D. H. Weiner & R. D. Nelson (Eds.), *Research in parapsychology 1986,* 82–85. Metuchen, NJ: Scarecrow Press.

OSIS, K., & MCCORMICK, D. (1980). Kinetic effects at the ostensible location of an out-of-body projection during perceptual testing. *Journal of the American Society for Psychical Research*, **74**, 319–329.

PALMER, J. (1971). Scoring in ESP tests as a function of belief in ESP: Part 1. The sheep-goat effect. *Journal of Parapsychology*, **65**, 373–408.

PALMER, J. (1977) Attitudes and personality traits in experimental ESP research. In B. B. Wolman (Ed.), *Handbook of parapsychology,* 175–201. New York: Van Nostrand Reinhold.

PALMER, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in parapsychological research 2,* 59–244. Jefferson, NC: McFarland.

PALMER, J. (1997). Correlates of ESP magnitude and direction in the PRL and RRC autoganzfeld data bases. *Proceedings of Presented Papers: The Parapsychological Association 43rd Annual Convention*, 283–298.

PERSINGER, M. A. (1989). Psi phenomena and temporal lobe activity: The geomagnetic factor. In L. A. Henkel & R. E. Berger (Eds.), *Research in parapsychology 1988,* 121–156. Metuchen, NJ: Scarecrow Press.

PERSINGER, M. A., & KRIPPNER, S. (1989). Dream ESP experiences and geomagnetic activity. *Journal of the American Society for Psychical Research*, **83**, 101–116.

PUTHOFF, H., & TARG, R. (1974). PK experiments with Uri Geller and Ingo Swann [Abstract]. In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.), *Research in parapsychology 1973,* 125–128. Metuchen, NJ: Scarecrow Press.

RADIN, D. I., MCALPINE, S., & CUNNINGHAM, S. (1994). Geomagnetism and psi in the ganzfeld. *Journal of the Society for Psychical Research*, **59**, 352–363.

ROE, C. A. (1996). Clients' influence in the selection of elements of a psychic reading. *Journal of Parapsychology*, **60**, 43–70.

ROE, C. A. (1998). Belief in the paranormal and attendance at psychic readings. *Journal of the American Society for Psychical Research*, **92**, 25–51.

RUBIN, L., & HONORTON, C. (1971). Separating the yins from the yangs: An experiment with the I Ching. *Proceedings of the Parapsychological Association*, **8**, 6–7.

SCHMEIDLER, G. R. (1973). PK effects upon continuously recorded temperature. *Journal of the American Society for Psychical Research*, **67**, 325–340.

SCHMEIDLER, G. R. (1988). *Parapsychology and psychology: Matches and mismatches.* Jefferson, NC: McFarland

SCHMEIDLER, G. R. (1994a). ESP experiments 1978–1992: The glass is half full. In S. Krippner (Ed.), *Advances in parapsychological research 7,* 104–197. Jefferson, NC: McFarland.

SCHMEIDLER, G. R. (1994b). PK: Recent research reports and a comparison with ESP. In S. Krippner (Ed.), *Advances in parapsychological research 7,* 198–237. Jefferson, NC: McFarland.

SCHMEIDLER, G. R., & MCCONNELL, R. A. (1958). *ESP and personality patterns.* New Haven, CT: Yale University Press.

SCHMIDT, H., & SCHLITZ, M. J. (1989). A large scale pilot PK experiment with prerecorded random events. In L. A. Henkel & R. E. Berger (Eds.), *Research in parapsychology 1988,* 6–10. Metuchen, NJ: Scarec+row Press.

SPIELBERGER, C. D. (1983). *State–Trait Anxiety Inventory for Adults: Form Y.* Redwood City, CA: Mind Garden.

SPOTTISWOODE, S. J. P. (1997). Geomagnetic fluctuations and free response anomalous cognition: A new understanding. *Journal of Parapsychology*, **61**, 3–12.

STANFORD, R. G. (1977). Experimental psychokinesis: A review from diverse perspectives. In B. B. Wolman (Ed.), *Handbook of parapsychology,* 324–381. New York: Van Nostrand Reinhold.

STORM, L., & THALBOURNE, M. A. (2000). A paradigm shift away from the ESP–PK dichotomy: The theory of psychopraxia. *Journal of Parapsychology,* **64**, 279–300.

TART, C. (1988). Geomagnetic effects on GESP: Two studies. *Journal of the American Society for Psychical Research,* **82**, 192–215.

THALBOURNE, M. A. (in press). *The common thread between ESP and PK.* New York: Parapsychology Foundation.

THALBOURNE, M. A., & DELIN, P. S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research,* **59**, 172–186.

TROSCIANKO, T., & BLACKMORE, S. J. (1983). Sheep-goat effect and illusion of control. In W. G. Roll, J. Beloff, & R. A. White (Eds.), *Research in parapsychology 1982,* 202–203. Metuchen, NJ: Scarecrow Press.

WIGGINS, J. S. (1989). Review of the Myers–Briggs Type Indicator. In J. C. Conoley & J. J. Kramer (Eds.), *Tenth mental measurements yearbook,* 536–538. Lincoln: University of Nebraska Press.

WILKINSON, H. P., & GAULD, A. (1993). Geomagnetism and anomalous experiences, 1868-1980. *Proceedings of the Society for Psychical Research,* **57**, 275-310.

Division of Psychology
University College Northampton
Boughton Green Road
Northampton NN2 7AL
United Kingdom
chris.roe@northampton.ac.uk

**Koestler Parapsychology Unit*
University of Edinburgh
7 George Square
Edinburgh EH 9JZ
United Kingdom
paul@moebius.psy.ed.ac.uk