APPENDIX 3:

THE SENDER AS A PK AGENT IN ESP STUDIES: THE EFFECTS OF AGENT AND TARGET SYSTEM LABILITY UPON PERFORMANCE AT A NOVEL PK TASK[17]

Nicola J. Holt & Chris A. Roe
Centre for the Study of Anomalous Psychological Processes
University College Northampton, Northampton, UK

ABSTRACT

Recent work has been concerned to evaluate whether the sender plays any active role in successful ganzfeld GESP experiments (e.g., Roe, Holt & Simmonds, 2003; Roe & Holt, in press). Roe, Holt & Simmonds (2003) used a random number generator (RNG) as a 'virtual receiver' in a ganzfeld, in an attempt to detect any sender effect. During the sending period descriptive statements were 'selected' from among a pool of 768 to give a 20-item 'RNG mentation' that may represent a more direct measure of any sender influence than the mentation of the 'human receiver'. A suggestive effect was obtained, with a 32.5% hit rate, when an independent judge (JW) used the 'virtual mentations' to select the target clip from three decoys ($Z = 1.48, p = .069$, 1-t). Roe and Holt (in press) sought to replicate this effect and further compared ganzfeld trials with no sender and standard ganzfeld trials. Support for the hypothesis that senders exerted some influence on the virtual receiver was obtained, psi success across two independent judges being higher in ganzfeld trials with a sender. JW obtained 42.1% hits in trials with a sender (SOR = 43, $Z = .821, p = .412$, 2-t) and 17.6% hits in trials with no sender (SOR = 47, $Z = -.868$, $p = .384$, 2-t), while RD (a newly recruited judge) obtained 26.3% hits in trials with a sender (SOR = 44, $Z = .616, p = .535$, 2-t) and 5.9% hits in trials with no sender (SOR = 46, $Z = -.651, p = .516$, 2-t). A third experiment in this series is presented here. The protocol was adapted in order to obviate the need for a human receiver. The focus for senders hence became the 'virtual receiver'. This displayed the statements to the sender as they were selected, as an analogue to hearing feedback from a human receiver in the ganzfeld. Senders could rate how well each statement corresponded with their sending experience. The lability of the target was manipulated (following Braud, 1981, 1994). Twenty-four statements were selected for each trial, from a pool of 416, eight by each of the following processes, which increased in lability: a random number table; a pseudo random process; and a live RNG. It was hypothesised that the greatest psi effect would be found with the most labile target. Further, drawing upon Stanford's conformance behaviour model (1978) it was hypothesised that senders with the most 'stable' trait characteristics would achieve higher psi hitting. Forty trials were conducted, the virtual mentations of which were rated by two independent judges (JW and LS). Significant psi hitting was not obtained in any of the randomness conditions, although there was a trend towards psi missing in the live condition for JW ($Z = -1.485, p = .069$, 1-t, $r = .235$) and a trend towards psi hitting in the pseudo condition for a newly recruited independent judge LS ($Z = 1.485$, $p = .069$, 1-t, $r = .235$). However, there was a significant interaction effect between target and sender lability, across both independent judges ($F_{4,74} = 4.959, p = .001$). The hypothesis that 'stable' senders would demonstrate higher psi hitting with the most labile target system was confirmed. Further, senders with high trait lability performed best with the most stable target system. This was interpreted as indicative of a reciprocal influence between labile and stable aspects of systems. Explanations for the overall lower psi outcome of this study were addressed in terms of the feedback potentially hindering motivation and the implications of direct rather than indirect intention, which was introduced in this study.

[17] We would like to gratefully acknowledge the financial support of the Bial Foundation which enabled us to conduct this study.

INTRODUCTION

It is not straightforward to determine whether a sender makes an active contribution to the success of ESP experiments. Although there does seem to be a subtle advantage for telepathy experiments over equivalent clairvoyant studies (see, e.g., Honorton, 1995, Ullman & Krippner with Vaughan, 1973), this might be explainable in simple psychological terms, such as the sharing of responsibility for failures (and successes) at a psi task and the calming effects of having a friend accompany one to an unusual and potentially anxiety provoking situation (Morris, Dalton, Delanoy & Watt, 1995). In earlier papers we described a novel method that promised to provide a more direct assessment of any sender contribution (Roe & Holt, in press; Roe, Holt & Simmonds, 2003) by introducing a 'virtual receiver' in the form of a random number generator (RNG) whose output was associated with a collection of 768 statements that had been coined to describe the clips in the target pool. During the normal sending period the RNG was continually sampled and a 'virtual mentation' produced that consisted of the 20 statements whose numbers had been selected most often. Independent judges then used the virtual mentations to rank order four video clips (the target clip and three decoys) in much the same way as the sender typically does in a conventional free response ESP study. The participants did not see the virtual mentation at any point.

In the first such study (Roe, Holt & Simmonds, 2003) the RNG was run in the receiver's room in an otherwise standard ganzfeld study. The sender was not asked to influence the RNG in any way, but to send information to their partner undergoing ganzfeld stimulation. An independent judge rated the virtual mentations according to their similarity to the target video clip and three decoys, as the 'human receiver' had done with their own personal mentation. The 'human receivers' rated the target clip most highly in 35% of the trials (where the MCE is 25%), a sum-of-ranks analysis for which was statistically significant ($Z = 1.77$, $p = .038$, 1-t). More relevant here, the statements selected by the 'virtual receiver' enabled the independent judge to correctly identify the target clip in 32.5% of the trials. While in the expected 'psi-hitting' direction, this was not statistically significant ($Z = 1.48$, $p = .069$, 1-t). This outcome was, however, considered to be encouraging and worthy of further investigation. Roe and Holt (in press) note that at r = .234, the effect size for the virtual receiver SOR analysis is comparable to previous work that used an RNG as a proxy receiver, where r = .257 (Roe, 1996). These effect sizes may be more comparable to those of DMILS studies, estimated at r = .33 (Braud & Schlitz, 1991) than that typically reported in REG-based psi research, which are estimated to be much lower at r = .00006 (Steinkamp, Boller & Bösch, 2002).

In a subsequent study, Roe and Holt (in press) attempted to replicate this finding and also compared the performance of the 'virtual receiver' in two conditions: ganzfeld trials with a sender and without a sender[18]. It was considered that the 'no sender' condition would act as a control against which to evaluate performance in the experimental condition. Two independent judges were also used in this replication. This was prompted by a concern raised by Roe, Holt and Simmonds (2003) that the selected statements could have been sufficiently flexible in how they might be interpreted by the independent judge to allow any above-chance scoring to be due to the judge's own psi, as they perhaps unconsciously selected the target and then contrived an interpretation of the statements after the fact that allowed it to be ranked first. The inter-rater reliability between the two judges employed by Roe and Holt (in press) was poor, with a Cohen's kappa of .202, suggesting that the mentations were indeed

[18] Aspects of this study focusing on the human receivers' performance in these conditions were reported by Roe, Sherwood & Holt (2003)

open to a range of interpretations. Across both the sender and no sender conditions the independent judge, JW, correctly selected the target in 30.6% of the trials, while a newly recruited judge, RD, had a below MCE hit rate of 16.7%. This appears to lend some support to the judge-ESP hypothesis. However, for both judges, performance was better in the experimental condition than the control condition. JW obtained 42.1% hits in trials with a sender (SOR = 43, $Z = .821$, $p = .412$, 2-t) and 17.6% hits in trials with no sender (SOR = 47, $Z = -.868$, $p = .384$, 2-t). While RD (a newly recruited judge) obtained 26.3% hits in trials with a sender (SOR = 44, $Z = .616$, $p = .535$, 2-t) and 5.9% hits in trials with no sender (SOR = 46, $Z = -.651$, $p = .516$, 2-t). The effect sizes for the sender trials were positive in this study, at r = .188 (JW) and r = .141 (RD), but smaller than found in earlier research. However, these results were considered of sufficient interest to warrant further study, being suggestive of a sender effect on the virtual receiver across both independent judges (although the differences between conditions were non-significant at $Z = 1.11$, $p = .134$ for JW and $Z = .824$, $p = .206$ for RD, this may be due in part to the low power of the study).

In planning the current replication, however, we were aware of the practical difficulties we had encountered in recruiting and scheduling sender-receiver pairs in these previous studies, and speculated that it might be worthwhile to brief participants more thoroughly about the RNG as virtual receiver and thereby obviate the need for a human receiver. We hoped that participants would be more focused on the nature of the psi task and that this might improve performance, particularly since by changing the participant's primary focus from a human sender to the RNG we were able to offer them statement by statement feedback. Following Roe (1996), the statements appeared as if somebody was typing them, even generating and correcting typing errors. It was envisioned that this would give the virtual receiver a 'human quality', and to enhance this impression the feedback was described to senders as analogous to hearing the mentation of the receiver in a standard ganzfeld experiment. Senders were able to rate these statements according to how closely they corresponded to their experience of sending. It was hoped that this level of involvement in the feedback process would outweigh any increase in skepticism on the part of participants at the idea of interacting with a computer rather than a human.

In addition to these alterations in protocol, the present study manipulated the source of randomness by which mentation statements were selected in order to explore the effects of target lability, following Braud (e.g. 1981, 1994) who has talked at length and persuasively about its importance in the action of PK. Braud & Schroeter (1983) have demonstrated how this can be manipulated practically by modifying the method by which random numbers are generated. They found that more labile selection methods (such as radioactivity-based random number generators) produced larger effect sizes than less labile ones (such as pseudorandom sources and random number tables), although this trend failed to achieve significance. One of us (Roe, 1996) attempted a simplified replication of this effect using a pseudopsychic paradigm in which participants rated the accuracy of statements generated by two methods that differed in lability. It was found that ratings for statements generated using an RNG source were higher (i.e. were considered more accurate) than those generated using random number tables, suggestively so by one measure and significantly so by another. Given the relative success of our previous studies in the current series, which had been derived in part from Braud and Schroeter (1983) and Roe (1996), we felt it worthwhile to extend the design to attempt to replicate this lability effect.

These lability effects might best be understood in terms of Stanford's conformance behaviour model (Stanford, 1978), which makes no distinction between ESP and PK but rather conceptualises psi as the 'conformance behaviour' of one system (e.g. an RNG or a human brain) to the needs of a 'disposed system' (e.g. a 'sender' in a psi experiment or a target image). The latter is characterised as a relatively ordered and stable system that is thus not predisposed to vary, whereas the conforming system is characterised as a relatively disordered or unconstrained system that is free to vary and thus may be able to change in ways that more closely conform to the disposed system. In terms of classical PK studies, Braud (1980, 1981) predicted that the likelihood and/or magnitude of a conformance behaviour effect would be proportional to the degree of lability, 'free variability' or 'capability for change' of the 'target system' and the degree of constraint, 'inertia' or structure of the PK agent. In this study we would therefore expect the greatest conformance behaviour with the most labile 'virtual receiver', but also with senders who present as most stable on personality and attitude measures. Braud, Shafer & Mulgrew (1983) defined cognitive lability as the degree of free variability in the mental processes of percipients. They assessed the cognitive lability of receivers by fluency of word associations and by 'perceptual lability' (the frequency of alternations between perceptual representations of a Necker cube), finding that only 'word fluency' was significantly correlated with subsequent psi scores (r = .39, $p < .05$, 2-t). In the present study, it was decided to widen the construct of lability/stability by considering lability of cognition, affect, experience and behaviour – trait factors tapping into the constraint/rigidity versus spontaneity that Stanford (1990) argued might moderate psi outcome.

Interacting with motivation and other cognitive and situational variables (e.g. belief in psi or reaction to the target clip), trait lability may lead to degrees of state lability of the PK agent while 'sending'. Braud (1981) stresses several state variables pertaining to structure that may impact on conformance behaviour: the degree of concentration, intention, analytical activity, effort or egocentric striving; and the relevance of thoughts/imagery to the task at hand. We should expect 'stable state' senders whose attention is focused on the target video clip, with clear, persistent and undistorted representations of it, to show a greater PK effect. Such aspects of the sender's state were to be assessed in a post experimental interview.

Study aims

Four predictions were hypothesised, using sum-of-target-ranks as the primary outcome measure. Predictions given below are directional but alpha levels for other, exploratory, analyses were conservatively kept as two-tailed.

1. RNG mentations will allow each independent judge to identify the target clip to a greater degree than expected by chance for each randomness condition.

2. Each independent judge's sum-of-ranks for live RNG statements (Live condition) will be lower than for pseudo random statements (Pseudo condition), which will be lower than for random number table statements (Table condition).

3. The participants' subjective ratings of the degree of correspondence between the mentation and their experiences and thoughts while sending information about the target video clip will be higher for live RNG statements than for pseudo random statements, which will be higher than for random number table statements

4. Less labile (i.e. more 'stable') individuals will demonstrate higher psi performance (Z-scores) on more labile systems.

Exploratory analyses were planned in order to consider the covariation between the independent judges' ratings (across the target lability conditions) and sender personality and attitude measures. Exploratory correlations were also planned between the participants' subjective correspondence ratings (across the target lability conditions) and the independent judges' ratings.

Further, a thematic analysis based on transcripts of post-trial interviews with the sender was planned in order to explore the nature of the sender's experience.

METHOD

Design

The present study is the third of a series that is intended to systematically explore the utility of using an RNG as a proxy receiver within a ganzfeld GESP protocol. However, in this study a simplified protocol obviated the need for a 'receiver', and there was a comparison of three target systems, which randomly selected descriptive statements with processes that differed in lability: live RNG; pseudo random; and random number table. The dependent variable for planned analyses is the sum-of-target-ranks awarded by independent judges; the dependent variable for exploratory analyses is the *Z*-score of target ratings.

Participants

An opportunity sampling method was used to draw 40 participants (mean age = 33.05 [range = 17 – 84], 18 males and 22 females). These consisted of friends and colleagues of the experimenter (N=12), students at UCN (N=10) and participants recruited from the wider community using posters and media appeals (N=18). 15 participants had participated in formal parapsychological experiments before, while 25 were novices. Lab personnel did not serve as participants. Nicola Holt conducted all trials in this study.

Participants may have had above average trait lability, scoring higher on openness-to-experience and neuroticism and lower on conscientiousness than the norm.

Apparatus and Materials

Details of the experimental suite have been described previously (Roe, Sherwood & Holt, 2003). In this study only one experimental room was required – the 'senders' room'. This study used an automated ganzfeld computer system developed by Dr Paul Stevens and written in Microsoft Visual Basic v5 that presented video material via the API for Media Player v7. Video clips are stored digitally as MPEG files, labelled 1a, 1b, 1c etc.

Part of the UCN target pool used in the last study was used for this study, consisting of 52 minute-long digital video clips that were drawn from commercial films to reflect a range of emotions and themes. Clips were arranged in 13 sets of 4 so that members of a set were as distinct as possible. Copies of the target pool are available on CD or DVD from the first author on request. Randomisation is achieved using the Visual Basic pseudo-random algorithm (rnd), seeded using the timer at the start of the program (RANDOMIZE TIMER). Once the 'Start' button has been pressed, the computer first selects a target set, then selects one of the four clips within that set.

The descriptor pool from which the RNG draws was different from the original study (Roe, Holt & Simmonds, 2003), as a different target pool was used. This consisted of eight statements for each of the 52 clips to give a total pool of 416. These statements were coined by the authors to describe the target set, but were intended to be essentially accurate yet not overly-specific (e.g., "I feel dreamy and trancelike" rather than "someone is hypnotising me") so that they were more characteristic of the kinds of descriptions given during ganzfeld stimulation, and also so that they could in principle help identify targets from other sets.

The mentation-generating program was written in QuickBasic v.1, and ran on an ACER Extensa 503T laptop running under Windows 98. The program was adapted from versions used in previous studies so as to be able to compare three different statements selection methods, using data generation methods as follows:

Random number table statements (the Table condition) for each participant were selected prior to commencement of the study using random number tables (Clark-Carter, 1997, Table X). An entry point to the list was determined using the RND function of a Casio fx-100 scientific calculator to give the row and the item along that row at which to begin the series. Reading from that entry point digits were considered in sets of threes and each value in the range 001 to 416 was taken to generate a single data file that was sufficiently long to cover all participants in the study. These data were arranged in a 24x40 array, with each row containing the pre-selected statement numbers for a single participant. The program checked whether a statement had already been selected for that participant for that condition, and where this was the case the next value in the series was used.

Statements using pseudorandom data (the Pseudo condition) were generated in real time using the INT(RND) command of Visual Basic to produce a value between 1 and 416. As for Table data, the program checked whether a statement had already been selected for that participant for that condition, and where this was the case a new value was generated.

For RNG data (the Live condition), the program sampled an Orion RNG v1.1 attached to a serial port. We required the RNG to generate numbers in the range 1-416, but this exceeds the 'natural' range of RNG outputs that runs from 0-255. Because of difficulties in combining more than one sample in a manner that ensured that all the possible outcomes were equally probable, we adopted a method in which for each selection the RNG was sampled 416 times, corresponding to the 416 statements. The iteration that generated the highest value became the selected statement (e.g., if only sample 117 returned the value 255 then statement 177 was selected). In the event of a tie, then the first sample to generate the joint-highest value was selected. Again the program checked whether a statement had already been selected for that participant for that condition, and where this was the case the process was repeated.

It was possible for the same statement to be selected and presented for more than one condition.

Materials

The standard UCN Participant Information Form (PIF) was adapted in this study. The resulting 15-item measure included questions concerning biographical and contact details (6-items); belief in PK (3-items); previous participation in parapsychological studies (2-items); practice of mental/physical disciplines (1-item); creativity (2-items); and self-perceived happiness (1-item). Copies of all in-house measures are available from the first author on request. Participants also completed a number of other measures pertaining to lability.

In study one (Roe, Holt & Simmonds, 2003) openness-to-experience was negatively correlated with psi-success (*rho* = -.266, *p* = .097, 2-t) and conscientiousness was positively associated with psi-success (*rho* = .212, *p* = .189, 2-t). Individuals who are open-to-experience are described as curious about both inner and outer worlds, and willing to entertain novel ideas and values, while conscientious individuals are conceived as 'directed', competent and striving towards achievement (Costa & McCrae, 1992). These trends were deemed of interest to the 'lability hypothesis'. Hence we used the NEO Five-Factor Inventory (NEO-FFI) (Costa & McCrae, 1992), a 60-item questionnaire with five subscales assessing: neuroticism, extraversion, openness to experience, agreeableness and conscientiousness. Each subscale has 12 items with a five-point Likert scale response format. As a measure of emotional instability and poor impulse control, neuroticism was included in the lability construct.

Persinger (1983) argues that all humans can be placed on a continuum of temporal lobe stability and that increased electrical lability of the temporal lobes may facilitate psi experiences (Persinger, 1989). He describes the person with temporal lobe lability as more likely to be impulsive, versatile, imaginative, emotionally sensitive, verbal, interested in philosophical questions and aesthetics and suggestible. The complex partial epileptic signs subscale of The Personal Philosophy Inventory (Persinger & Makarec, 1987) was used in this study, which consists of 16-items pertaining to temporal lobe lability (e.g. visions, hearing inner voices, intense sensations of smells without an obvious source, sense of noesis, perceptual aberrations, bodily vibrations, and dissociation from 'reality') with a dichotomous ('yes'/'no') response scale.

A measure of mood lability was also used, which was developed to screen for bipolar disorders (Akiskal et al. 1995). This consisted of 2-items with a 7-point Likert response scale ranging from 'not at all' to 'very much so': "My mood often changes from happiness to sadness, without knowing why" and "I have frequent ups and downs in mood, with and without apparent cause".

Braud (1981) equates the 'novelty generation' involved in the creative process with lability. Hence, aspects of creativity were included in the lability construct. Holt, Delanoy and Roe (2004) found that rather than cognitive, personality or domain based components of creativity, only one creativity component, termed 'intrapersonal awareness' was significantly correlated with the incidence of spontaneous parapsychological experiences (r = .45, p = .000001). This component may be considered in terms of emotional and cognitive lability and was composed of scores on two measures:

The Emotional Creativity Inventory (Averill, 1999), a 30-item scale with a 5-point Likert response scale that measures three facets of emotional experience: preparedness; novelty; authenticity and effectiveness.

The Creative Cognition Inventory (unpublished measure by Holt), a 29-item scale with a 5-point Likert response scale, assessing the use of different cognitive styles in the creative process, with seven sub-scales: heightened internal awareness; intuition and inspiration; linear, goal-directed cognition; playful, absorbed cognition; the use of analogy; and oneiric cognition (e.g. ideas arising in dreams).

Each trial was followed by a semi-structured interview in which participants were asked about their subjective impressions of the success of the trial, their experience of participating and the type of sending strategies used.

Procedure

Potential participants were sent an information sheet illustrated with photographs that described most aspects of the study. This provided a rationale for the 'sender as PK agent' paradigm, and outlined the stages of the experimental procedure. Prior to the trial, participants completed a battery of measures. Participants were greeted on arrival and escorted to a reception room that had been specially prepared with comfortable chairs, a coffee table, rugs, paintings and curtains so as to make participants feel as comfortable and relaxed as possible prior to the trial. The experimenter encouraged an informal and positive atmosphere, discussing the procedure and answering any questions that arose while sharing refreshments. Participants were informed of previous ganzfeld studies that had used an RNG as a proxy receiver. They were told that this study was an extension of this research, without the added complications of a 'human receiver', and were encouraged to focus on enjoying and getting involved in the task. Participants were not initially aware of the different target lability conditions in this study. Participants were then shown the 'sender's room' and their role was explained again and they were made familiar with the setting and equipment.

The sender was seated in a comfortable office chair. They were shown the laptop computer and made aware that randomly-selected statements would appear on this screen. They were shown a number keyboard with which they were to rate each statement after it had appeared. The sender typed in their name and date into the laptop, after which a set of instructions appeared for them to read. These informed the sender that when they were ready to begin the experiment, they should press the space bar on the keyboard. The sender was shown an adjacent computer monitor on which the target video clip would be displayed. They were shown how to adjust the volume of the video clip on the speakers, and how to pause the video clip if they wished. They were also asked if the lighting was at a comfortable level and shown the dimmer switch for the room's lighting in case they wished to alter it during the trial. When the sender was ready to begin and had no more questions, the experimenter initiated the programme that selected the video clip to begin playing in one minute's time, wished the sender good luck and waited in the adjacent reception room for the trial to finish. Any information that the sender might need during the trial was clearly presented on posters on the wall of the sender's room in case they felt anxiety about remembering any of it.

The trial commenced with the randomly-selected video clip being played to the sender. This one-minute long clip was set to play 14 times, with a five second pause between each repetition. The sender was advised to watch the clip once in order to become familiar with it, and then to initiate the selection of random statements on the laptop. When they did this, a statement would appear on the screen, letter by letter, as if being typed, during which process it would sometimes make a typing error and go back to correct it. Once the statement was complete a question mark would appear at the bottom of the screen to cue the sender to rate how similar the statement was to their experience (using a 9-point scale, where 1 = not at all similar and 9 = highly similar). When participants had typed in a correspondence rating, another statement would begin to appear after a delay of a few seconds. The sender was encouraged to rate the statements according to the entire content of their experience rather than just literal associations with the target clip, for example feelings of tiredness, seemingly tangential thoughts, etc. In total twenty-four statements were rated by each sender, eight of each target system. For each participant the statements were drawn in a consistent sequence from the three target lability systems, and this order was counterbalanced across participants. When all 24 statements had been presented and rated a message was displayed, stating that the trial had been completed and thanking the sender for taking part in the study and asking them to inform the experimenter in the adjacent room that they had finished.

The experimenter and the sender chatted informally and then the sender's permission was asked to record the subsequent discussion. All participants agreed to this. Together the sender and experimenter reviewed the statements that had been selected during the trial, looking at a list of these and their correspondence ratings that constituted the saved record of the trial. They discussed any correspondences that seemed of interest to the sender and their experiences of sending, which led on to a semi-structured interview focusing on the sending strategies that they used. The participant then had the opportunity to ask the experimenter any questions about the study and was asked if they would like to know more about the aims of the study, in which case they were told about the randomness conditions and the lability hypotheses.

After completion of all trials in the series, two judges who were otherwise uninvolved with the study[19] independently rated the four clips in each set according to the degree to which they reflected the content of the mentation. This was done separately for each of the target lability conditions: RNG, Pseudo and Table. The independent judges were blind to this manipulation and were simply asked to perform the judging for 120 trials, each based on 8 mentation statements. Each trial was given a new and independent code so that it was not possible to identify the sequence in which the trials occurred or any pattern between them without cross-referencing.

RESULTS

The source of randomness and senders' subjective correspondence ratings

The feedback of statements meant that the sender could evaluate the degree of correspondence between each statement and their own experience. It was expected that statements in the Live condition, as the most labile source, would receive higher correspondence ratings from senders, however this was not the case. Table 1 shows the mean scores for each target lability condition. Since correspondence ratings were on a 9-point scale, with 8 statements in each condition, this gives total correspondence scores in the range 9-72. The values in Table 1 equate to an average rating per statement of only 3.5. The subjective correspondence ratings are marginally higher for the Table condition and lowest for the Live condition. A Friedman test (a non-parametric omnibus test of within-participant differences) showed that there was no significant difference between the subjective correspondence ratings of statements generated by the three target systems ($\chi^2 = .517$, 2df, $p = .772$).

TABLE 1:

SENDER'S SUBJECTIVE RATINGS OF THE DEGREE OF CORRESPONDANCE BETWEEN THEIR EXPERIENCE AND THE RANDOM STATEMENTS GENERATED BY THE THREE TARGET SYSTEMS

	Table	Pseudo	Live
M	29.70	29.03	28.20
SD	11.84	10.87	11.67
N	40	40	40

[19] We are grateful to Jacqui Wilson and Louie Savva for serving as independent judges.

Inter-rater reliability in ranking clips of the independent judges

For each condition, of each trial, the independent judges compared the 8 statements to the four video clips for that trial, giving each clip a confidence rating (based on similarity between the clip and the mentation) between 0 and 99%. These were converted into target rankings, from '1' if they rated the target clip highest to 4 if they rated it lowest. We were concerned to assess the degree to which the independent judges' ratings would be determined by the mentation, with the alternative being that there was sufficient scope for 'interpretation' that judging may reduce to an elaborate forced choice ESP task for the independent judge. The target rankings of the two independent judges were calculated for each target lability condition, with Cohen's Kappas that were considered poor in each case: $\kappa = .299$ ($p = .001$) for the Table condition; $\kappa = .096$ ($p = .287$) for the Pseudo condition; and $\kappa = .036$ ($p = .688$) for the Live condition. As a consequence of this low inter-judge reliability it was decided not to combine their rankings but to consider them separately.

Target lability and psi performance

The ranks allocated to target clips by the independent judges based on the three randomly generated mentations (Live, Pseudo and Table) are reported in Table 2.

For the ratings by JW the direct hit rates for all target systems are below mean chance expectation (MCE = 25%). For both the mentations generated by the random number table and by the pseudo random process the direct hit rate is 22.5% and for the live RNG the direct hit rate is 12.5%. The sum-of-ranks do not differ significantly from chance expectation for any of the target lability conditions. JW worked on our two previous studies, giving above MCE direct hit rates of 42.1% ($Z = .821$, $p = .412$, 2-t, $r = .188$) and 32.5% ($Z = 1.485$, $p = .069$, 1-t, $r = .235$) in ganzfeld conditions with a sender, where statements were generated by a live RNG in the receiver's room. In this study, for the live RNG condition there is a trend towards psi missing, with an effect size comparable to that of the earlier studies ($Z = 1.485$, $p = .138$, 2-t, $r = .235$). In the other conditions however, the effect sizes are not comparable, being $r = .011$ in the pseudo condition, and $r = .034$ in the table condition.

For the ratings by LS, a new independent judge, the direct hit rates for all target systems are above mean chance expectation. For both the mentations generated by the random number table and by the live random process the direct hit rate is 27.5% ($r = .012$) and for the pseudo random process the direct hit rate is 32.5%. The sum-of-ranks does not differ significantly from chance expectation for the latter, but again has an effect size commensurate with the above ($Z = -1.485$, $p = .138$, 2-t, $r = .235$).

The ratings for the independent judges have different outcomes, with a trend towards psi missing (12.5% hit rate) in the live condition for JW, and a trend towards psi hitting in the pseudo random condition for LS (32.5% hit rate). Both JW and LS have ratings close to MCE for the Table condition, and it was in this condition that they had the greatest agreement ($\kappa = .299$, $p = .001$). It may be that the independent judges are picking up on different information in the mentations based on individual differences in judging strategy or conditions of judging.

TABLE 2:

A COMPARISON OF TARGET RANK FREQUENCIES FOR MENTATIONS BY TARGET SYSTEM (N = 40)

	Target System	Rank						p (2-tail)
		1	2	3	4	SOR	Z	
Independent Judge JW	Table	9 (22.5%)	11 (27.5%)	9 (22.5%)	11 (27.5%)	102	.212	.832
	Pseudo	9 (22.5%)	12 (30%)	10 (25%)	9 (22.5%)	99	-.071	.943
	Live	5 (12.5%)	12 (30%)	10 (25%)	13 (32.5%)	111	1.485	.138
Independent Judge LS	Table	11 (27.5%)	11 (27.5%)	8 (20%)	10 (25%)	97	-.076	.940
	Pseudo	13 (32.5%)	13 (32.5%)	6 (15%)	8 (20%)	89	-1.485	.138
	Live	11 (27.5%)	11 (27.5%)	8 (20%)	10 (25%)	97	-.076	.940

Covariation of psi performance in each target lability condition with the senders' subjective correspondence ratings

For each lability condition, the ratings of the independent judges were converted into Z-scores, a standardised measure of distance and direction of the target rating from the mean rating for the four targets in each trial. This was deemed preferable to using the simple rank, since it is more sensitive and allows for greater variance across participants, which is essential when considering covariation.

These were compared to the participants' correspondence ratings for each condition, i.e. the cumulative ratings (0-9) for all 8 statements for that condition. In addition to a lack of difference in senders' subjective impressions of accuracy of each statement across the target lability conditions, there was no significant correlation between senders' correspondence ratings and the independent judges' ratings in each condition. Table 3 shows that these all have effect sizes below *rho* = .151. It appears that the senders and judges saw different links and associations between the statements and the target clip.

Interaction between lability of the sender and lability of the target system

A composite score of sender lability was created, by summing scores on all the trait lability measures that were selected on a theoretical basis in order to cover lability of behaviour, emotions, cognition and experience. These components all formed significant correlations with the composite measure in the expected direction, as follows: the use of non-linear forms of cognition (*rho* = .828, *p* = .0000001, 1-t); emotional creativity (*rho* = .743, *p* = .0000001, 1-t); temporal lobe lability (*rho* = .647, *p* = .00003, 1-t); openness-to experience (*rho* = .436, *p* = .0001); neuroticism (*rho* = .306, *p* = .028, 1-t); the use of linear cognition (*rho* = -.331, *p* = .018, 1-t); and conscientiousness (*rho* = -.296, *p* = .032, 1-t). This composite lability variable had a Kaiser-Meyer-Olkin statistic of .737 suggesting that its components shared sufficient variance to represent an underlying construct. Scores on the composite lability measure approximated the normal curve (Kolmogorov-Smirnov statistic = .113, df = 40, p = .200), ranging between 121 and 270, with a mean score of 211.5.

A mixed 3x3x2 ANOVA was conducted, with psi performance (*Z-scores*) as the dependent variable. Factor 1 was 'target system lability', within-subjects with three levels: Table, Pseudo and Live. Factor 2 was 'sender lability', and was a between-subjects measure with three levels: low (N = 13); medium (N = 14); and high (N = 13) scores on the composite lability measure (based on a 3-way split). Factor 3 was 'judge', between-subjects, level one being JW and level 2 being LS.

TABLE 3:

SPEARMAN RHO CORRELATIONS BETWEEN THE SENDER'S SUBJECTIVE CORRESPONDENCE RATINGS AND PSI PERFROMANCE BASED ON THE INDEPENDENT JUDGE'S RATINGS FOR EACH TARGET SYSTEM (AND 2-TAILED PROBABILITIES)

		Independent Judge JW			Independent Judge LS		
		Table	Pseudo	Live	Table	Pseudo	Live
Participant's correspondence ratings	Table	.063 (.698)	-	-	.098 (.549)		
	Pseudo	-	-.061 (.708)	-		.151 (.386)	
	Live	-	-	.131 (.421)			-.087 (.592)

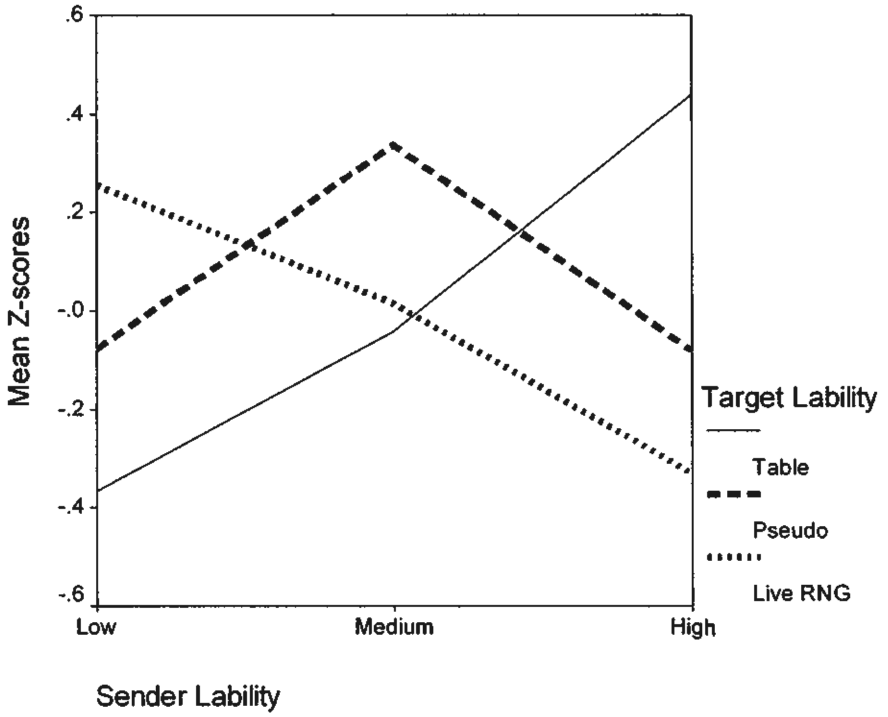
There were no significant main effects, neither the degree of target lability ($F_{2,74}$ = .169, p = .845), nor the lability of the sender ($F_{2,37}$ = .559, p = .651) significantly impacted upon psi-success. Neither was there a significant difference in overall scoring between the independent judges, although there was a trend for LS to obtain higher Z-scores overall ($F_{1,37}$ = 2.056, p = .160). The independent judges appeared to obtain a similar pattern of scoring, in that there was no difference between their ratings across either the three target lability conditions ($F_{2,74}$ = .690, p = .418) or the three sender lability conditions ($F_{2,74}$ = .894, p = .418). However, there was a significant interaction effect between the lability of the target system and the lability of the sender ($F_{4,74}$ = 4.959, p = .001)[20]. Further, a similar pattern appeared to emerge for both independent judges, there being no significant difference between the target and sender lability interactions of JW and LS ($F_{4,74}$ = 1.227, p = .307). The form of the target x sender lability interaction is displayed in Figure 1.

As hypothesized, stable (low lability) senders performed at the highest level with statements generated by the most labile system (the live RNG) (with hit rates of 23% [JW] and 46% [LS]), while highly labile senders performed at the lowest level with statements generated by the live RNG (with hit rates of just 8% for both JW and LS). Conversely, highly labile senders performed at the highest level with statements generated by the most stable system

[20] This is significant when corrected for multiple analyses with the Bonferroni method where the criterion level for significance is .003 (.05/18).

(the random number table) (with hit rates of 46% [JW] and 54% [LS]), while 'stable' senders performed at the lowest level in this condition (with hit rates of 0% [JW] and 8% [LS]). It appears that 'medium labiles' had more psi hitting in the pseudo random condition (the 'medium' labile system), with hit rates of 36% (JW) and 50% (LS).

FIGURE 1:
INTERACTION BETWEEN TARGET AND SENDER LABILITY ON PSI PERFORMANCE



Conducting post-hoc tests revealed that the interaction effect consisted of the following significant differences: high lability senders had significantly higher *Z*-scores in the Table condition than the Live condition (*t[12]* = 3.056, *p* = .010, 2-t) and stable senders had significantly higher *Z*-scores in the live RNG condition than the table condition (*t[12]* = -2.495, *p* = .028, 2-t).

Interpreting these data in consideration of earlier sum-of-ranks analyses reminds us that there was a trend towards psi hitting in the pseudo condition in LS's judging. This appears to be accounted for by medium labiles performing well. Indeed, a sum-of-ranks analysis for medium labiles in the Pseudo condition alone is significant (*Z* = -2.512, *p* = .012, 2-t, *r* = .671). The earlier sum-of-ranks analyses also suggested a trend towards psi missing in the Live condition for JW. This seems to be accounted for by both medium (*Z* = 1.794, *p* = .073, 2-t, *r* = .479) and high labiles (*Z* = 1.240, *p* = .215, 2-t, *r* = .343) psi missing.

Thematic analysis of sender strategy

A thematic analysis was conducted on the post-experimental interviews. As this has led to a wealth of data, this will be considered in more detail in a subsequent paper. The main distinction that emerged concerning sender strategy concerned intention – 'consciously willing'. 32% of participants described consciously attempting to influence the random process at some point during the trial, focusing on particular ideas/words or emotions and 'willing' them to appear within the statements, e.g. "I actually thought about it … pull pull pull… erm, so yes, I was literally going please say something … yes I was telling the computer to do something!". 68% of participants described their activity as 'just doing', simply watching the clip and rating the statements, often with a range of embedded techniques, such as trying to notice something different in the clip each time, free associating or focusing on memories and emotional reactions to the clip, for example: "I totally forgot I was supposed to be sending ... I forgot I was a sender, I was just sitting here wondering whether they would be the same ... that was it, I wasn't actually trying to send whatsoever r... I forgot".

All participants reported taking part in the study to be a positive experience. However, motivation to do well in the sending task was low, with only 28% of participants affirming this. Others either were simply 'curious' or 'just seeing what would happen', or deliberately stepped back from 'trying too hard' as a strategy. Participants reported varying degrees of absorption versus distraction during the PK task. 63% found the video clip engaging and easy to get involved with, for example "yes, you do get so absorbed that you don't actually realising you are doing it". Others had difficulties paying attention to the video clip. For 21% this was because they found the dual nature of the task difficult, switching between the screen playing the video clip and the screen showing the statements, sometimes to the extent that the content of the statements themselves were influencing their thoughts, rather than the video clip. For example: "I did feel sometimes that I was looking at the screen here with the writing on instead of looking at the picture... and sometimes that was manipulating what I was thinking". 21% found the statements not only distracting but frustrating or demotivating, for example: "I got a bit downcast, thinking, this is not working".

Discussion

Contrary to expectation this study did not find that the virtual mentations enabled independent judges to identify the target clip at a level significantly greater than chance expectation in any of the randomness conditions. This may be interpreted as an indication that no psi was involved in the processes studied, or that any psi effects are small and emerge in a complex system. There was insufficient agreement in terms of target ranking by the independent judges to consider their ratings collectively (Cohen's Kappas being considered poor in each case: $\kappa = .299$, .096, and .036). Effect sizes ranged from r = .011 to .235, with the latter being commensurate with those from previous studies (Roe, Holt & Simmonds, 2003; Roe & Holt, in press), although not all were in the predicted direction. There was a trend towards psi missing in the live condition for JW ($Z = -1.485$, $p = .069$, 1-t, $r = .235$) and a trend towards psi hitting in the pseudo condition for LS ($Z = 1.485$, $p = .069$, 1-t, $r = .235$). This suggests that the judges saw different points of similarity between the mentations and the target and decoy clips. It may be that the statements are too ambiguous, allowing much freedom of interpretation. It would be interesting to compare agreement ratings here with other studies that have used more than one independent judge.

The senders' own ratings of correspondences between the statements and their experience of sending, however idiosyncratic, were not significantly different across the three randomness conditions ($\chi^2 = .517$, 2df, $p = .772$). Following on from the work of Fox (2000) which considered the thought processes of senders in a ganzfeld study, it is intended to study these ratings further, in order to explore factors pertaining to lability, such as the types of associations made (e.g. literal or metaphorical) and the extent to which they deviated from the content of the target clip. For example, the following link between falling bottles (target clip) and blood (a statement) seemed highly relevant to one participant, an association that would be easily missed by an independent judge: "I really agreed with [that statement] because ... you know that bit where all the bottles fall ... I thought they were like little molecules ... and it was really weird, it just felt like little ... like haemoglobin." Clearly, then, senders may be impressed with correspondences that arise from personal associations that a receiver is typically going to be unaware of. When the associations, rather than direct references to the target clip, predominate (as suggested by Fox's innovative analysis) then it clearly makes the independent judge's (and receiver's) task all the more difficult even where psi might have occurred.

It was not found that psi-success increased as lability of the randomness condition increased (Table; Pseudo; Live RNG). ANOVAs reiterated this, with no main effect for target lability on psi outcome. Neither was there a main effect for trait lability of the sender. However, there was a significant interaction between senders' trait lability and target lability ($F_{74,4} = 4.959$, $p = .001$). The hypothesis that 'stable' senders would demonstrate higher psi hitting with the most labile target system was confirmed. However, the interaction appeared to be complex, confirming a speculated prediction of a 'mirror' effect, where senders with high trait lability performed best with the most stable target system. Also contributing to the interaction was a non-significant 'intermediary effect' where medium labile senders performed best in the pseudo random condition. These results concur with Braud's lability hypotheses and his extension of Stanford's conformance behaviour model, suggesting that in a PK task 'order' may be introduced into randomnicity under certain optimal conditions. However, these findings emphasise a bi-directional process between the labile and stable aspects of a system.

Proceeding with caution, these results must be interpreted with the trend towards psi-missing of high labiles with the most labile target in mind – any possible psi effect in this interaction may arise in this condition. Braud (1981) suggests that two labile systems may both conform to a more stable influence, which may be for example, an experimenter. Such situational variables may have contributed to shifting the direction of the effect compared to earlier studies.

Overall poorer psi performance compared with the previous studies (Roe, Holt & Simmonds, 2003; Roe & Holt, in press) may be explained by participants having been more labile than the norm, with high mean scores on openness-to-experience and neuroticism and low mean scores on conscientiousness, making it more difficult for the live condition to succeed. Further, it may be that lability of the independent judges interacts with this process, noticing some associations and not others.

The trend towards psi-missing rather than psi-hitting (in the live RNG condition) may also have been a product of the overt nature of this task, which differed in nature from the covert nature of previous studies, where a live RNG ran in the background of a ganzfeld study. Here, the intention was direct (i.e. to influence the statements) rather than indirect (i.e. to influence a human receiver).

The lability interactions uncovered in this study are considered to be of sufficient interest to warrant further exploration. In the next planned study we intend to explore the lability hypothesis further, and to select participants with this in mind, in order to represent a wide range of lability. Further, due to comments from participants about distractibility on the task and frustration or de-motivation due to lack of perceived correspondences between the statements and their experience, we will investigate more directly the effects of having direct feedback, by manipulating this variable. In the next study a 2x2 design may help understand the effects of having participants know about the virtual receiver (direct versus indirect strategies) and comparing end-of-session feedback with statement-by-statement 'real time' feedback, while revisiting this study's interesting findings by keeping the three data generation methods.

REFERENCES

Akiskal, H., Maser, J., Zeller, P., Endicott, J., Coryell, W., Keller, Warshaw, M., Clayton, P., & Goodwin, F. (1995). Switching from "unipolar" to bipolar II: An 11-year prospective study of clinical and temperamental predictors in 559 patients. *Archives of General Psychiatry,* **52**, 114–123.

Averill, J. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality,* **67**, 331-371.

Braud, W. (1980). Lability and inertia in conformance behavior. *Journal of the American Society for Psychical Research*, **74**, 297-318.

Braud, W. (1981). Lability and inertia in psychic functioning. In B. Shapin & L. Coly (Eds) *Concepts and theories in parapsychology*. New York: Parapsychology Foundation, inc. pp. 1-36.

Braud, W. (1994). The role of mind in the physical world: A psychologist's view. *European Journal of Parapsychology*, **10**, 66-77.

Braud, W., & Schroeter, W. (1983). Psi tests with Algernon, a computer oracle. In Roll, W., Beloff, J., & White, R. (Eds.) *Research in Parapsychology 1982* (pp 163-165). Metuchen, NJ: Scarecrow Press

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

Braud, W., Shafer, D. & Mulgrew, J. (1983). Psi functioning and assessed cognitive lability. *Journal of the American Society for Psychical Research*, **77**, 195-208.

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove, East Sussex: Psychology Press.

Costa, P., & McCrae, R. (1992). *Professional manual: Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (FFI)*. Florida: Psychological Assessment Resources, Inc.

Fox, J. (2000). A systems approach to the investigation of telepathy. *The 24th International Conference of the Society for Psychical Research*, University College Northampton.

Holt, N., Delanoy, D., & Roe, C. (2004). Creativity, subjective paranormal experiences and altered states of consciousness. *Proceedings of presented papers: The Parapsychological Association 47th Annual Convention,* 433-436.

Honorton, C. (1995). Impact of the sender in ganzfeld communication: Meta-analysis and power estimates. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 132-140.

Morris. R., Dalton, K., Delanoy, D., & Watt, C. (1995). Comparison of the sender / no sender condition in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 244-259.

Persinger, M. (1983). Religious and mystical experiences as artifacts of temporal lobe function: a general hypothesis. *Perceptual and Motor Skills* **57**, 1255–1262.

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

Persinger, M., & Makarec, K. (1987). Temporal lobe epileptic signs and correlative behaviors in normal populations. *The Journal of General Psychology*, ***114***, 179–195

Roe, C. (1996). Clients influence in the selection of elements of a psychic reading. *Journal of Parapsychology*, **60**, 43-70.

Roe, C., & Holt, N. (in press). A further consideration of the sender a PK agent in ganzfeld ESP studies. *Journal of Parapsychology.*

Roe, C., Holt, N., & Simmonds, C. (2003). Considering the sender as a PK agent in ganzfeld ESP studies. *Journal of Parapsychology*, **67**, 129-145.

Roe, C., & Sherwood, S., & Holt, N. (2003). Interpersonal psi: Exploring the role of the sender in Ganzfeld GESP tasks. *Proceedings of Presented Papers: The Parapsychological Association 46th Annual Convention*, 237-256.

Stanford, R. (1978). Toward reinterpreting psi events. *Journal of the American Society for Psychical Research*, **72**, 197-214.

Stanford, R. (1990). An experimentally testable model for spontaneous psi events: A review of related evidence and concepts from parapsychology and other sciences. In S. Krippner (Ed.), *Advances in parapsychological research, Vol. 6* (pp. 54-167). Jefferson, NC: McFarland.

Steinkamp, F., Boller, E., & Bösch, H. (2002). Experiments examining the possibility of human intention interacting with random number generators: A preliminary meta-analysis. In C. Watt (Ed.), *Proceedings of presented papers: The Parapsychological Association 45th Annual Convention*, 256–272.

Ullman, M., & Krippner, S., with Vaughan, A. (1973), *Dream telepathy: Experiments in nocturnal ESP*. Jefferson, NC: McFarland.