Roe, C.A., Martin, H., & Drennan, S. (2014). Effects of participant and target system lability upon PK performance using an I Ching task. *Journal of the Society for Psychical Research*, 74, 219-238. ISSN: 0037-9751

EFFECTS OF PARTICIPANT AND TARGET SYSTEM LABILITY UPON PK PERFORMANCE USING AN I CHING TASK

By CHRIS A. ROE, HANNAH MARTIN & SOPHIE DRENNAN[1]

ABSTRACT

Relatively few parapsychological experiments investigating micro-PK effects have been designed to consider psychological or individual differences factors, and those variables that have been considered have been subject to too few replications to give a clear indication of which persons may perform best under which conditions. Previous research by the first author discovered and replicated an interaction effect between an individual differences factor, participant lability, and a situational factor, target system lability. The present study was designed to conceptually replicate that finding using a novel task so as to control for possible artifacts. An alternative task was built around the I Ching divination procedure, which it was felt retained important characteristics of being personally relevant for the participant and intuitively straightforward to understand. An opportunity sample of 34 participants completed a measure of lability and decided upon a personal question that the I Ching could help with. Participants were run individually and completed a Q-sort of all 64 hexagram descriptions based on their applicability to their question. Once completed they cast three hexagrams using a computer based program that used a live random number generator (Live), the pseudorandom function of the computer (Pseudo) and a predetermined list of random numbers derived from published tables (Table). The Q-sort positions were used to rate the applicability of the selected hexagrams. Although the general pattern of performance was in line with prediction, with the highest average ratings awarded to hexagrams selected by the most labile Live method, next highest for the moderately labile Pseudo method and worst ratings for the most stabile Table method, the mean shifts were small and nonsignificant. Similarly, although the highest overall performance was achieved by the most labile participant group, an intermediate level of performance was recorded by the intermediate group and worst performance was by the stabile group, the modest differences were not significant. Therefore, despite the pattern of performance being superficially similar to that reported in previous studies, this experiment was not able to replicate the interaction between participant and target system lability. Possible causes for this failure to replicate are considered.

[1] We should like to thank The Bial Foundation for their kind support of this project (104/08). An earlier version of this paper was presented at the combined Parapsychological Association 53rd Annual Convention & The Incorporated Society for Psychical Research 32nd Annual Convention, Paris, 2010.

INTRODUCTION

Interest in dice throwing as a means of testing for PK declined with the advent of more sophisticated approaches that exploit advances in technology. Among these advances has been the use of Random Number Generators (RNGs) as alternate sources of randomness, which have allowed for study designs that reduce the opportunities for participant fraud and recording errors, and most importantly—in our view—have enabled feedback to be provided to the participant in a form that is more personally meaningful to them, for example, by creating engaging game environments (e.g., Berger, Schechter & Honorton, 1986; Broughton & Perlstrom, 1986, 1992). PK-RNG has a reasonable track record of success. Radin and Nelson (1989) reviewed 152 references describing 832 discrete studies by 68 different investigators and found that while control trials had a combined z score of zero, experimental trials gave a z score of 3.25 ($p < .0005$); 54,000 nonsignificant 'file drawer' studies are needed to eliminate the reported effect. One might reasonably conclude from this that the occurrence of the anomaly has been established.

However, researchers have been less successful in identifying characteristic features of the anomaly that might give some insight as to mechanism (or indeed simply reduce the likelihood that the observed statistical deviations are a result of some as-yet unknown methodological artifact). Most PK research seems to have been conducted by parapsychologists with professional training in physics and an interest in exploring the implications of parapsychological phenomena for problem solution and theory development in that discipline (Roe, 2001). This has had consequences for the nature of the experimental designs adopted and thus for our understanding of the role (if any) of participant variables. For example, Rush (1986 p. 62) bemoaned that

> most parapsychological experimentation has sought primarily to demonstrate that significant manifestations can be obtained with varied experimental techniques and procedures. Only a few experiments have been designed to explore the PK process by controlled tests of psychological or physiological variables

And according to Irwin (1999, p. 140)

> little research has been directed to the psychological dimensions of PK performance. Much of the current work, despite the welcome improvements in precision and control, still is concerned with the existence of PK and the forms it may take.

This lack of interest in psychological variables is evident from Radin and Ferrari's (1991) meta-analysis, which reports that although the number of participants ranged from 1 to 393 the median was only 3, with a number of active researchers content to conduct studies in which they were the sole experimenter-participant. Such studies can tell us little about psychological factors from a nomothetic perspective. Although some researchers have made some initial explorations of the effects of personality (e.g., Schmidt & Schlitz, 1989), belief (e.g., Gissurarson, 1997), Anxiety (e.g., Broughton & Perlstrom (1992), prior experience (e.g., Gissurarson & Morris, 1991), volitional strategy (e.g. Houtkooper, 2000), and arousal (e.g. Braud, 1985), there is a great deal of scope for further, more systematic investigation.

Work by the first author has focused on psychological factors that affect PK performance and has identified the construct of lability as particularly promising. This research adopted a protocol initially intended to test the suggestion that the sender in a conventional telepathy experiment might serve as a PK agent (Roe & Holt, 2005; Roe Holt & Simmonds, 2003). An RNG was placed in the receiver's room during an otherwise standard ganzfeld experiment. The RNG acted as a 'virtual receiver' by producing data that were used to select statements from among a large set descriptors that would be more or less applicable to the target video clip that the sender was attempting to send. Any PK effect might be used to select those descriptors that most accurately described the clip so as to enable an independent judge to identify it when presented alongside decoy clips, much as the judge would rate the human receiver's mentation.

In the first such study (Roe, Holt & Simmonds, 2003) a 32.5% hit rate was obtained (MCE = 25%; $z = 1.485$, $p = .07$). In a conceptual replication, Roe and Holt (2005) compared sender and no sender trials to see whether the original finding might be attributable to the performance of the judge rather than any sender effect — a likely explanation if the effect persisted on trials where in fact there was no sender. Some support was obtained for the hypothesis that senders would exert an influence on the virtual receiver, as psi success (using two independent judges, JW and RD) was higher in trials with a sender than those without (JW sender trials gave 42.1% hits, $z = .821$, $p = .41$, no sender trials gave 17.6% hits, $z = -.868$, $p = .38$; RD sender trials gave 26.3%, $z = .616$, $p = .54$, no sender trials gave 5.9% hits, $z = -.651$, $p = .52$). Although ultimately nonsignificant, the outcomes of these studies were considered sufficiently encouraging to warrant further research.

Holt and Roe (2006) simplified the protocol by removing the human receiver and explicitly briefed participants that their task was to influence the RNG so as to have it select accurate descriptions of a clip they were simultaneously watching. In addition, the lability of the target system was manipulated, with some statements being selected by a relatively labile method (a live RNG), some by an intermediate method (the computer's own pseudo random process), and others by a relatively stabile method (a random number table). It was hypothesized that the greatest psi effect would be found with the most labile target system (following Braud, 1981, 1994), and that senders with the most 'stabile' trait characteristics (as assessed using a composite that included measures of cognitive, emotional, behavioral and perceptual lability) would perform best at a PK task. Further, drawing upon Stanford's (1978) conformance behavior model, it was hypothesized that there would be an interaction between participant and target system lability in their effects upon psi performance. This expected interaction effect between target and sender lability was found, $F_{4,37} = 9.96$, $p = .001$, as senders with lower trait lability achieved higher psi scores in the highest labile target condition and vice versa. Roe and Holt (2006) confirmed this pattern of performance in a replication study with a further 40 participants; a mixed 3x3 ANOVA found a significant interaction between target lability and sender lability, $F_{4,74} = 2.747$, $p = .03$.

Taken together, these findings suggest that the approach adopted to test for PK provided an effective way of capturing anomalous effects, giving effect sizes that seem to be some orders of magnitude greater than is typical in PK studies (see, e.g.,

Roe Holt & Simmonds, 2003), which we interpret as being at least partly due to our attempts to generate PK tasks that are engaging for participants and which have some personal meaning for them. However, the interaction effect between sender and target system lability needs to be confirmed, ideally through independent replication by other researchers so as to help rule out explanations in terms of methodological artifact. Alternatively, further replications by the current author that involved a new task utilizing a different computer program (particularly in the sense that how the RNG output is translated into the participant's perceived outcome) could also assuage concerns over some possible methodological artifact.

It was important that this new task still gave rise to an outcome that would be intuitively meaningful to the participant while being sufficiently different from previous tasks to control for artifact. With this in mind, an I Ching-based 'divination' task was considered the most promising candidate since similar RNG-based 'readings' had been successful previously (Roe, 1996) and have proven to be extremely popular as part of classroom demonstrations, suggesting that they would continue to be engaging here. Among divination methods, the I Ching seemed most suitable because it already relies on the interpretation of a set of random processes, as described in the following section.

The I Ching is a method of divination that is said to have originated in China (Thalbourne, Delin, Barlow & Steen, 1992-3) and consists of a sequence of six binary outcomes to give 64 unique permutations. Typically the enquirer poses a question upon which they seek advice and then casts yarrow stalks or coins as a random method by which to generate a particular hexagram. The hexagram is made up of six lines that can either be yang lines (——————) or yin lines (——— ———) or 'changing-line' versions of each. The statement(s) associated with that hexagram are then interpreted as an answer to the posed question. This divination method makes some appeal as a means of testing for PK, since it involves an ostensibly random system whose results are believed to reflect the agent's needs or intentions, and which gives rise to outcomes that are amenable to statistical analysis.

The first attempt to incorporate the I Ching into a psi task was by Rubin and Honorton (1971), who recruited 40 participants — most of whom were naïve regarding the I Ching — to register their attitude to ESP using a 10-point response scale. They then had to think of a question, the answer to which had great personal meaning to them. While they reflected on this question they threw 6 pennies from a cup. This was repeated until the hexagram was completed. Participants were subsequently given two hexagrams to read; the correct one and a randomly selected control. Each hexagram was rated on a 10-point scale according to the degree to which it was relevant to the question asked. Overall, results were reported to be nonsignificant (although no statistics are given) but the 24 sheep did score significantly higher than the 13 goats (1.75 versus –3.23, $t = 2.22$, $p < .05$).

Thalbourne *et al*. (1992-3) attempted to replicate this finding with 53 participants, but this time the method used just 3 coins thrown 6 times (see Thalbourne, 1994, for an extended description of the method). In week 1, participants completed a measure of attitude to I Ching and the Australian Sheep-Goat Scale (ASGS) then wrote their personal question on a slip of paper which was sealed in an envelope. They then cast the coins and recorded the numbers of heads and tails for each of six throws. The

experimenter used this record to produce a hexagram and a control hexagram that was generated by converting every yin line to a yang line and vice versa. This process was done mechanically (in the sense that readings were taken verbatim from a book rather than interpreted subjectively) so that it may not have been a design flaw to have the hexagrams produced by someone who was not blind as to which was genuine and which the control. In week 2 participants were presented with their two readings and rated them on a 10-point scale for relevance. This delay before feedback may not have been psi-conducive and also introduced an opportunity for participant cheating, since they could have memorized their outcome and have looked up the correct interpretation in preparation for the judgment they knew they would have to make in week 2. No-one admitted to looking up their interpretation when questioned about this, but this is not a sufficient safeguard. These shortcomings could have been addressed by the use of an automated system. On average the actual interpretation was rated as more relevant than the control, but this difference was not significant ($t = 0.38$, p not reported). They did find a positive but non-significant correlation between this measure and sheep-goat scores ($r = .16$, $p = .13$), but a stronger prediction using attitude to I Ching scores ($r = .24$, $p < .05$).

Storm and Thalbourne (1998-99) conducted a conceptual replication in which participants were presented with the 64 descriptors that corresponded to the possible hexagram outcomes and were asked to select the 16 (25%) that most applied to them. They then cast coins to 'select' one hexagram and a hit was registered if this corresponded to one of the 16 they had previously selected as most appropriate. The likelihood of a hit by chance was therefore 25%. Storm and Thalbourne (1998-99) reported a hit rate of 32%, which is suggestively higher than chance expectation of 25% (binomial $p = .067$), and they did find a significant correlation between hitting and scores on a measure of transliminality ($r = .27$, $p = .01$). This latter finding may have been compromised, however, by some participants having received feedback about their performance on the I Ching task before they completed the transliminality measure.

The relative success of this study is remarkable given the very crude method by which success of outcome was measured. A more sensitive measure could be produced by having participants rate the accuracy of each statement relative to the others, for example by employing the Q-sort method (Stainton Rogers, 1995). This requires participants to place statements in a pre-specified normal distribution with free slots (a choice of 64 hexagrams fits very neatly into such a shape, which naturally has 64 slots; see Figure 1). In this way the accuracy of any selected hexagram can be coded in terms of its actual position in the distribution and hence as parametric data. This allows us, for example, to talk meaningfully about single trials being 'significant', since there is a 1/64 chance of the selected hexagram being the most preferred statement (+7), a 3/64 likelihood ($p = .047$) of it being rated +6 or greater, and so on. It also allows for greater variance in outcome to compare against variation in other measures, such as personality and attitude variables.

Storm and Thalbourne (2001) extended their research with a larger sample (N = 107) using a method that was similar to Storm and Thalbourne (1998-99) except that no feedback was given to any participants until after personality measures had been completed, thus avoiding earlier criticism. They reported an improved hit rate of

35% (binomial p = .015), but failed to replicate the positive correlation with transliminality (r = .01, p = .48). However, participants were again responsible for manually casting their own hexagrams. Although participants may not be aware of the meaning of outcomes in such a manner as to be able to consciously affect the interpretation, it is still not ideal to have a manual divination process like this, which unnecessarily allows for the introduction of unconscious biases — J. B. and Louisa Rhine recognized at an early stage the problems inherent with PK dice studies in which participants threw the dice from their hand or even from a cup, and they quickly shifted to more automated procedures (cf. Rhine, 1970).

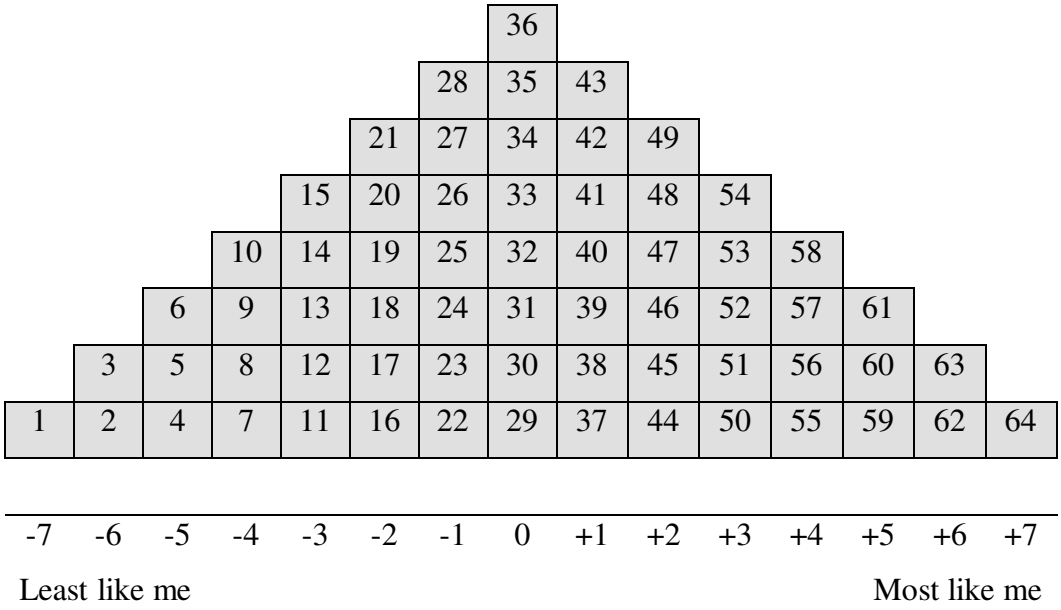


Figure 1. Q-sort distribution for 64 alternatives

To conclude, the few tests of PK that have incorporated an I Ching task have been reasonably successful in eliciting above chance performance that seems to co-vary meaningfully with individual differences measures. This may be because the method engenders the properties of psychological meaning and personal relevance for the participant that were identified earlier in relation to the first author's own work as especially psi conducive yet notably absent from most current PK research designs. It is surprising that other research teams have not attempted independent replications. Some may have been deterred by the rather loose controls in place, with manual casting of hexagrams being reminiscent of early Rhinean PK research that now might be thought to epitomize poor control. However, these methodological weaknesses can be addressed by developing a fully automated procedure whose randomness is derived from widely accepted sources such as commercial RNGs.

Aims of the Present Study

The present study was intended to be a conceptual replication of earlier findings of an interaction between the effects of participant lability and target system lability upon performance at a PK task (Holt & Roe, 2006; Roe & Holt, 2006). A new task was proposed that built on the work of Rubin and Honorton (1971) and Thalbourne

and colleagues in using an I Ching task. Methodological weaknesses in those studies due to manual casting of the I Ching and opportunity for human error in interpreting and feeding back readings would be overcome here by adopting an automated design in which I Ching castings would be initiated by the participant but conducted by computer (using the three lability methods used previously by Holt & Roe, 2006 and Roe & Holt, 2006). Criticisms concerning the crudity of the outcome measure were addressed here by having participants Q-sort all 64 possible hexagram readings prior to using the I Ching program. Thus, the following predictions were made:

H_1: Mean rating for the hexagram selected by the Live (RNG) selection method will be significantly greater than zero.

H_2: Mean rating for the hexagram selected by the Pseudo (pseudorandom algorithm) selection method will be significantly greater than zero.

H_3: Mean rating for the hexagram selected by the Table (predetermined values derived from random number tables) selection method will be significantly greater than zero.

H_4: There will be a main effect of participant lability group upon ratings for selected hexagrams.

H_5: There will be a main effect of target system lability upon ratings for selected hexagrams.

H_6: There will be an interaction between participant lability group and target system lability upon ratings for selected hexagrams.

METHOD

Design

This experiment utilized a 3x3 mixed ANOVA design to consider the effects of participant lability (categorized as high, moderate or low) and lability of hexagram selection method (high, medium or low using a live RNG, pseudorandom algorithm and random number tables respectively). The dependent variable was the Q-sort rating given to the selected statement by the participant prior to its 'selection' in the divination casting. The nature of the Q-sort method ensures that ‘sample’ data are drawn from a normally distributed set of ‘population’ scores.

Participants

An opportunity sample of 34 participants completed the experiment (12 males, 22 females; mean age = 25.8, *SD* = 11.1, range = 18-56). Participants were not screened for prior ability or experience of divination or PK, but were recruited based on their interest in the project and their willingness to commit over an hour to participation. However, subsequently collected questionnaire data indicates that a majority (67%) described themselves as having no PK ability and 71% felt they would not be able to demonstrate PK effects under the conditions of the study; the effects of belief and expectation will therefore be considered in the Results section.

Questionnaire measures

A participant information form (PIF) comprised standard briefing instructions and 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).

Previous studies of lability by Holt and Roe (2006) and Roe and Holt (2006) have included a large battery of measures to gauge lability, but this was deemed too demanding for current purposes and likely to discourage participation or lead to partial completion of the measures. Therefore the decision was taken to streamline the measure of participant lability by restricting it to only those variables that had correlated with the composite score in both previous studies with $r > .4$. Thus measures of mood lability, neuroticism, linear cognition and conscientiousness were excluded here. The measures that were included were

- The Creative Cognition Inventory (unpublished measure by Holt), a 29-item scale with a five-point Likert format assessing the use of different cognitive styles in the creative process, with two main subscales: the use of linear versus nonlinear cognition.
- The Emotional Creativity Inventory (Averill, 1999), a 30-item scale with a five-point Likert format that measures three facets of emotional experience: preparedness; novelty; authenticity and effectiveness. This measure has acceptable internal consistency (alpha = .89; Gutbezahl & Averill, 1996)
- The complex partial epileptic signs subscale of the Personal Philosophy Inventory (Persinger & Makarec, 1987), 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.
- Openness to Experience was measured using Goldberg (1999). This measure is derived from the international personality item pool (IPIP: Goldberg *et al.*, 2006), and has subscale alphas that range from .77 to .86 (Goldberg, 1999). It was chosen for use here because it is a public domain measure that was intended to represent the domain constructs of the NEO personality inventory (Buchanan, Johnson & Goldberg, 2005). Correlations between the IPIP and NEO scales for the six facets of the openness to experience dimension range from .70-.80 (Goldberg, 1999), suggesting that these instruments measure the same personality dimension.

Apparatus & software

A hexagram-generating program was written in QuickBasic v.1 (copy available from the first author on request), and ran on an ACER Extensa 503T laptop running under Windows 98. The program simulated a hexagram casting procedure based on the coin method used by Thalbourne and colleagues (Storm & Thalbourne, 1998-9; Thalbourne *et al.*, 1992-3). In this method three coins are tossed to determine each

line of the six-line hexagram such that different outcomes give a yin line (▬ ▬), a yang line (▬▬▬), or a changing-line version of each of these. The procedure is repeated to build the hexagram line by line from its base, taking onto account the consequences of changing lines (see Yang & Sandifer, 2003, chapter 5; Ritsema & Sabbadini, 2005, pp. 12-18). The emerging hexagram is illustrated on-screen (see Figure 2). Once the sequence of lines is completed the program identifies the associated hexagram and presents its interpretation as feedback.

Figure 2. Screen presentation of I Ching hexagram casting

This procedure is repeated three times to give I Ching castings that use each of the three different target selection methods, with the order of conditions randomized across participants. The target selection methods operate as follows:

- In the Table condition, random number table values 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. Data were stored as values in the range 0-255 to mirror the other target selection methods. Data were treated as 'tails' for values between 0 and 127, and as 'heads' for values between 128 and 255. Data were considered in sets of threes to simulate throwing three coins, and the hexagram line associated with the outcome was identified. This procedure was repeated for each of the six lines that make up the hexagram.
- In the Pseudo condition, pseudorandom data were generated in real time using the INT(RND) command to produce a value between 0 and 255. Thereafter the procedure was as for Table data.

- In the Live condition the program sampled an Orion RNG v1.1 attached to a serial port. The 'natural' range of RNG outputs runs from 0-255. Thereafter the procedure was as for Table data.

It was possible for the same hexagram to be generated and presented for more than one condition.

Hexagram readings & Q-sort sheet

Previous studies have tended to provide single-word descriptors as outcomes from I Ching castings. In order to better simulate the conditions of earlier studies on lability and to provide participants with more meaningful feedback, it was decided to generate single sentence descriptors for each of the 64 hexagrams. These were coined by the first author and were derived from a number of popular guides to interpreting the I Ching (Riseman, 1980; Ritsema & Sabbadini, 2005; Yang & Sandifer, 2003; Wilhelm, 1951/2003). Care was taken to make the descriptions as authentic as possible, ensuring that the descriptors captured the essence of the slightly varying interpretations from different sources, while simultaneously attempting to make the statements as distinct from one another as possible, so as to facilitate their evaluation by participants. Example statements include "*T'ai (Peace)*: It is better to be truthful than to say what you think others want to hear", "*Yu (Enthusiasm)*: It's fine to indulge yourself sometimes, but this isn't particularly productive for you if it isn't balanced by hard work", "*Kuei Mei (The Marrying Maiden)*: An offer may seem too good to be true because it is. All may not be as it appears".

All the hexagram statements were reproduced on small laminated cards so that they would be sufficiently robust for repeated use during Q-sorting. A Q-sort grid was printed on A2 paper (16.5" × 23.4") and also laminated for durability.

Procedure

Participants were recruited by opportunity sampling; they were given a briefing that described the main practical elements of the study and stressed that the intention was to test the validity of a new means of conducting I Ching readings rather than to test their own ability or personality. Participants were advised that the practical aspect of the experiment would take approximately one hour to complete and that they were welcome to decline without penalty should they be unable to devote this amount of time to the study. Those who consented to participate were provided with a copy of the PIF to complete in their own time before the I Ching consultation and were asked to spend some time beforehand thinking of an appropriate personal question that could form the basis of their reading — it was made clear that they would not need to declare what that question was but simply keep it in mind during the I Ching casting.

At their scheduled session the participant was shown the Q-grid and the experimenter explained how they should bring to mind the question that they would like to ask the I Ching and then rate the 64 descriptors against one another for their applicability to that question, with more applicable statements being placed to the right of less applicable ones (as described by Stainton Rogers, 1995; see illustration

in Figure 3). The final arrangement should place one descriptor in each of the 64 cells of the grid, such that one item was identified as most applicable (rated +7), two were rated as next most applicable (+6), three items as next most applicable (+5), and so on. Once it was clear that the participant understood the task, the experimenter withdrew to allow them to complete the task in their own time and without observation. When they were satisfied that they had arranged the descriptors in the most suitable order, the participant informed the experimenter, who made a record of the number positions of all 64 statements.

Figure 3. *Mock-up of the Q-sort interaction, illustrating the grid and descriptor items*

The participant was then introduced to the I Ching program. The experimenter explained the basic principle of casting a hexagram and the participant was positioned in front of the screen, instructed to think about their personal question, and when the moment felt right to begin the casting by pressing the space bar. The participant watched as the hexagram was built and a statement identified. The second and third castings were similarly initiated by the participant when they were ready. On completion of the third hexagram, the participant was debriefed as to the nature of the study. They were provided with contact details that allowed them to request feedback on the study outcome or, if they so wished, anonymously to withdraw their data at a later date.

Ethical considerations

There was a possibility that participants might take the advice provided by the I Ching too seriously and this could potentially have adverse consequences. We took care here to frame the experiment as a test of this method of casting the I Ching so that there was no prior assumption that its advice might be sound, and we also ensured that participants were reminded at the end of the I Ching session that it should be treated as a fun exercise rather than any serious attempt to resolve personal issues. Care was also taken in the construction of the hexagram readings to ensure that none was extremely positive or negative, so as to reduce their possible

impact upon people's post-experiment behaviour. Otherwise, the study adhered to the BPS's Code of Conduct and Ethical Guidelines (available at http://www.bps.org.uk/the-society/code-of-conduct/code-of-conduct_home.cfm), in particular in ensuring that no deception was involved in the study, that all data were recorded and kept in an anonymous form, and providing participants with a mechanism to withdraw from the study should they so wish.

RESULTS

I Ching data from one participant were corrupted, reducing the sample for analysis to 33. Participants' ratings for the hexagram statements are generated in such a way that they are drawn from a normally distributed population of scores. It can be seen from kurtosis and skew data given in Table 1 that all three distributions are reasonably normal given the modest sample size, and so parametric testing was deemed appropriate.

	Mean	**SD**	**Skewness (Std. Error)**	**Kurtosis (Std. Error)**
Table source	-.30	3.53	.11 (.41)	-.62 (.80)
Pseudo source	.36	3.50	-.39 (.41)	-.51 (.80)
Live source	.50	3.03	-.23 (.41)	-.37 (.81)

Table 1. Descriptive statistics for participant ratings of the hexagrams subsequently selected for them by the three target selection methods

Participants rated all 64 hexagram statements on a rating scale that runs from -7 (least applicable) to +7 (most applicable) using a Q sort distribution. Thus if the statement selected by the computer program by each method were determined by chance alone then we would expect the average rating to approximate zero, whereas if the selected statements were rated as more accurate than the average statement then this mean value would be positive, and if it were less applicable than average among these statements then its rating here would be negative. We can see in Table 1 that there is a modest positive shift in ratings for hexagrams generated using the Pseudo and Live sources but a modest negative shift for the Table source. When analyzed by 1-sample t-test against the null value of zero, however, we find that these deviations are not significant (for Live, $t[32] = .93$, $p = .36$; for Pseudo, $t[32] = .60$, $p = .56$; for Table, $t[32] = -.49$, $p = .63$), failing to support H_1, H_2 and H_3.

However, the primary prediction was not that there would be an overall effect but rather that there would be internal effects that reflected interactions between participant and situational variables. Table 2 presents details of PK performance broken down by participant lability and target system lability.

We can see that the low lability group achieved the lowest overall average rating for their selected hexagrams. There is an increase in ratings as we move to the moderate and then the high lability groups, but these differences are modest given

the much larger variations within-groups as indicated by the standard deviation values; the differences in scores between lability groups are not significant ($F_{2,29}$ = .099, p = .906).

Greatest success was achieved overall with the most labile target system (Live), intermediate performance was with the Pseudo selection system and worst performance was with the relatively stable Table target selection, broadly in line with prediction. However, again these differences are quite small, ranging from only +.50 to -.30 against a background of much larger within-condition variability (standard deviations range from 3.03 to 3.53); the differences in scores between target selection methods are not significant ($F_{2,58}$ = .571, p = .568).

		Lability group			
		Low	moderate	high	Total
Target selection source	Table	-1.09	-.55	.73	-.30
		(3.62)	(3.05)	(3.95)	(3.53)
	Pseudo	-.18	.55	.73	.36
		(4.38)	(3.53)	(2.65)	(3.50)
	Live	1.27	.30	-.09	.50
		(2.24)	(3.30)	(3.53)	(3.03)
	Average	.00	.27	.45	.24
		(2.24)	(2.63)	(2.37)	(2.34)

Table 2. Mean (and standard deviation) of participant ratings by participant lability group and target selection source

Our principal objective in this study was to replicate earlier findings that suggested an interaction between the two variables thus far considered. Data that pertain to this interaction are illustrated in Figure 4. Interestingly, this graph depicts a pattern of performance that is broadly similar to that reported in earlier studies (Figures from those papers are reproduced here for comparison purposes). For example, in all three cases scores for the Live condition show a decline from the Low lability group through the moderate lability group and give the lowest scores for the high lability group. In contrast, performance in the Table condition is worst for the low lability group, here and in Holt & Roe (2006) improves through the moderate lability group, and in all three cases gives the best performance in that condition among the high lability group. No clear predictions were made regarding the intermediate Pseudo condition, but the pattern here is reminiscent of Holt & Roe (2006). However, when subject to a 2-way analysis of variance, this interaction is not close to significance ($F_{4,58}$ = .896, p = .473), and so H_6 must be rejected.

Finally, we noted earlier that in this sample many participants reported that although they were open to the possibility of PK they did not believe that they had PK ability nor would they be capable of producing PK effects in the context of this experiment. As a *post hoc* exploration it was decided to reanalyse the data to see if

this amounted to a self-fulfilling prophesy. Participants were divided into 'sheep' and 'goats' and the effects of belief on performance were tested using a 2x3 mixed ANOVA to incorporate the different target systems. Results indicated that there was no main effect of prior belief or expectation ($F_{1,30} = .062$, $p = .804$), nor was there an interaction between belief and target condition ($F_{2,60} = .086$, $p = .918$).

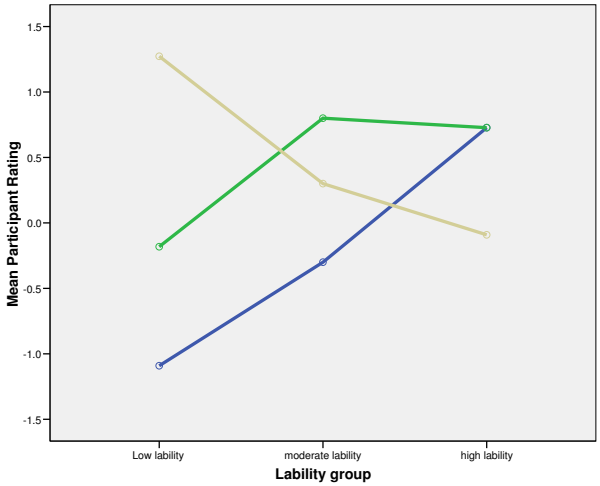
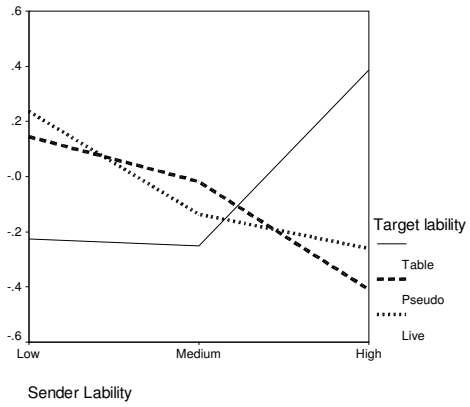
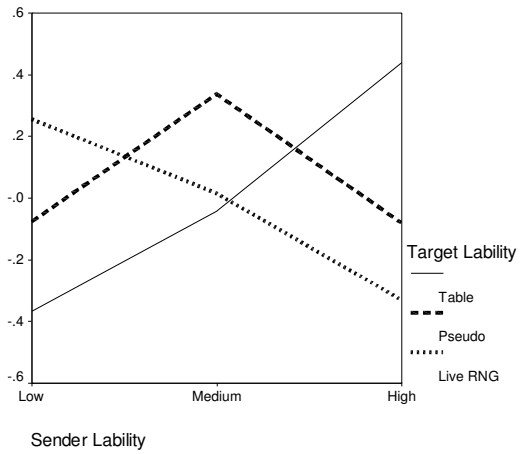


Figure 4. Interaction between target source and participant lability group from the present study



Interaction figure from Holt & Roe (2006)



Interaction figure from Roe & Holt (2006)

DISCUSSION

Although mean scores attributed to the selected hexagram exhibited the predicted pattern in that the highest average ratings were awarded to hexagrams selected by the most labile Live method, were next highest for the moderately labile Pseudo method and gave an average rating worse than the null value of zero for the most stabile Table method, these mean shifts were very small (at +.50, +.36 and -.30 respectively) and, given the wide variation in ratings across individuals, did not

come close to significance. Similarly, as expected the highest overall performance was achieved by the most labile participant group, an intermediate level of performance was recorded by the intermediate group and worst performance was by the stabile group, but these modest differences are not close to significance (+.45, +.27 and .00 respectively), and one should conclude that this study did not provide any evidence of PK.

However, in those earlier studies the lack of significance in tests for main effects was attributable to a significant interaction which suggested that two effects were at work that cancelled each other when combined. That is, in both studies, stabile participants showed a marked preference for the most labile (Live) target system compared with the least labile (Table), whereas labile participants showed a marked preference for the least labile (Table) target system compared with the most labile (Live), which was evident in significant interaction analyses. This pattern is consistent with Stanford's (1978) conformance behavior model, which was one of the inspirations of the original study's design. In Figure 4 we saw that stabile participants again show best performance in the Live condition (indeed the best performance overall) and worst performance in the Table condition (the worst performance overall); for labile participants (joint) best performance is with the most stabile Table condition and worst for the Live condition, although the differences here are much smaller. Despite appearances, however, the interaction between participant lability and target system lability was not close to significance, so that H_6 was rejected and we must conclude that this experiment has failed to conceptually replicate the lability interaction effect (Holt & Roe, 2006; Roe & Holt, 2006).

It may be that those earlier findings are spurious, although one referee commented on an earlier paper (Roe & Holt, 2006) that the reported effect sizes were 'remarkably strong'— indeed, throughout this series of studies, the effect sizes seem to be some orders of magnitude greater than is typical in PK studies (see, e.g., Roe, Holt & Simmonds, 2003), which we interpret as being at least partly due to our attempts to generate PK tasks that are engaging for participants and which have some personal meaning for them. It seems unlikely that the discrepancy in outcome here is due to this factor: efforts were again made to design an experiment that participants would find engaging and which they felt invested in, and informal feedback from participants here suggests that that was the case, as many reported that they had enjoyed the task and had felt motivated to find out something about themselves (either through the Q-sort or the I Ching itself) even if it should not be taken too seriously, much like a newspaper horoscope reading.

On reflection, working with unselected participants could be regarded as a strategy that is doomed to failure, particularly where — as in this case — we subsequently discover that a majority (67%) described themselves as having no PK ability and even more (71%) felt that they would not be able to demonstrate PK effects under the conditions of the study. Given that some researchers have previously found that paranormal belief predicts PK performance (e.g., Gissurarson & Morris, 1991; von Lucadou, 1987; Morris, Dumughn, Gentles & Grice, 1993), one could clearly argue for the application of some screening measure. However, it might not be productive to base that screening on participants' belief or expectation of success: when the effects of participant belief were considered in a *post hoc*

analysis here they indicated that this was not a significant factor in accounting for variance in performance. Alternatively, one could recruit all participants to a pilot test and invite back for formal testing only those who showed some success in that initial screening. Such an approach could be quite resource expensive, however, particularly if PK performance shows itself to be not particularly reliable (Boller & Bösch's, 2000, test-retest figures for micro-PK ranged from .269 to -.045, which is far from adequate).

Notwithstanding these comments, there is clearly a lot of variance in scores here that is not accountable for in terms of measured variables and this has contributed a lot of 'noise' to the dataset that could potentially obscure rather more subtle effects among the measured variables. Rather than screening participants, a more efficient solution could be to conduct a much larger replication in which a multivariate approach is made possible (following Schmeidler's, 1988, recommendation) and this might incorporate a range of variables including those that have been associated with PK performance previously but could not be included in a project of this scope, such as state and trait anxiety (Roe, Davey & Stevens, 2003), personality (e.g., Feeling-Perceiving dimensions of the Myers-Briggs Type Indicator: Schmidt & Schlitz, 1989), prior experience (Gissurarson & Morris, 1991), volitional strategy (Houtkooper, 2000), and arousal (Braud, 1985).

Another avenue for future work would be to consider the experimenter-participant interaction (Roe, Davey & Stevens, 2006). In this study all participants were recruited by opportunity sample and were known to the experimenter so there was a degree of psychological closeness (or at least familiarity) but over a session period of approximately an hour there is plenty of opportunity for smaller variations in experimenter-participant rapport to have an impact upon performance. It would be worthwhile to explore this formally by having experimenter and participant complete assessments of that rapport and of their optimism concerning the session (as incorporated previously by Roe, Davey & Stevens, 2006, and Sherwood, Roe, Holt & Wilson, 2005).

Since divination methods such as the I Ching and Tarot are often used periodically by members of the public (see Blackmore, 1983; Ivtzan & French, 2004) it may also be informative to conduct a longitudinal study with a smaller number of participants to consider variations in quality of 'advice' in relation to, for example, perceived need, mood variables, or environmental factors such as geomagnetic activity and local sidereal time (see Braud & Dennis, 1989).

Finally, Storm and Thalbourne (2001) have rightly noted that successful I Ching studies may be interpreted as evidence of the action of PK on the method of divination but equally could be interpreted in terms of precognition of a future outcome affecting ratings given at the present time. It would be informative to see whether these two explanations can lead to different predictions. Earlier work by the first author (e.g., Roe, Stevens & Davey, 2003) adopted a protocol that allowed ESP and PK conditions to be combined in a manner that allowed for 'disguised' trials (i.e. ESP trials that were presented as PK trials and vice versa), and perhaps something similar could be devised here to distinguish between PK and precognition. Indeed, one might suggest that the current design effectively does so, since it is difficult to imagine how success in the Table condition could be

understood as being brought about by PK, so that if similar patterns of performance are found here as for the Live condition (which is more obviously amenable to putative PK effects) then this may suggest that all the effects are due to some form of ESP combined with demand characteristics. Should these two conditions give markedly different patterns of performance then this might indicate that they are brought about by different mechanisms (ESP in the case of the Table condition, PK in the case of the Live condition), particularly if those differing patterns were consistent with the wider literature in ESP and PK. However, the reported effects would need to be more robust before these speculations can be taken too seriously.

Further work with the I Ching task could incorporate some methodological improvements. It would be important to test the RNG output for randomness during control periods and to gather sufficient data to be able to assess the output distribution for experimental periods (there were too few data points in the present study for this to give a meaningful assessment). It would also be useful to get feedback from participants concerning the hexagram statements used here, since it is possible that some statements are more likely to be generally applicable than others in a manner similar to the Barnum Effect (see Roe, 1995, for a consideration of the relevance of the Barnum Effect to psychic readings). Finally, there is a need to better articulate the construct of lability as an individual difference measure. This has previously shown some promise as a moderating factor in PK performance, but at present the construct is poorly specified and has not been subject to any psychometric evaluation.

Centre for the Study of Anomalous Psychological Processes
Psychology Division
University of Northampton, UK

REFERENCES

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

Berger, R.E., Schechter, E.I. & Honorton, C. (1986). A Preliminary Review of Performance across Three Computer Psi Games. In Weiner, D.H. & Radin, D.I. (Eds), *Research in Parapsychology 1985* (pp 1-3). Metuchen, NJ: Scarecrow Press.

Blackmore, S. (1983). Divination with Tarot Cards: An empirical study. *JSPR*. **52**, 97-101.

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*, pp. 380-382.

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.G. (1985). ESP, PK and sympathetic nervous system activity. *Parapsychology Review*, **16(2)**, 6-16.

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

Braud, W.G. & Dennis, S.P. (1989). Geophysical variables and behavior: LVIII. Autonomic activity, hemolysis, and biological psychokinesis: Possible relationships with geomagnetic field activity. *Perceptual & Motor Skills*, **68**, 1243-1254.

Broughton, R.S. & Perlstrom, J.R. (1986). PK Experiments with a competitive computer game. *JP*, **50**, 193-211.

Broughton, R.S. & Perlstrom, J.R. (1992). PK in a competitive computer game. *JP*, **56**, 291-306.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment*, **21**, 116-128.

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

Daniels, M. (1998). Transpersonal psychology and the paranormal. *Transpersonal Psychology Review*, **2(3)**, 17-31.

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

Gissurarson, L.R., & Morris, R.L. (1991). Examination of six questionnaires as predictors of psychokinesis performance. *JP*, **55**, 119–146.

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds), *Personality psychology in Europe (vol 7)*. Tilburg, Netherlands: Tilburg University Press. 7-28.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C. R., & Gough, H.G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, **40**, 84-96.

Gutbezahl, J., & Averill, J. R. (1996). Individual differences in emotional creativity as manifested in words and pictures. *Creativity Research Journal*, **9**, 327-337.

Holt, N., & Roe, C.A. (2006). The sender as a PK agent in ESP studies: The Effects of agent and target system lability upon performance at a novel PK task. *JP*, **70**, 69-90.

Houtkooper, J. (2000). Mind-machine interaction: Two phases of replication in Giessen & Freiburg. *Proceedings of presented papers: The Parapsychological Association 43rd Annual Convention*, p. 419.

Irwin, H.J., & Watt, C.A. (2007). *An introduction to parapsychology (5th ed.)*. Jefferson, NC: McFarland.

Ivtzan, I. & C. French, C.C. (2004). Testing the validity of tarot cards: Can we distinguish between a real and a control reading? *Proceedings of presented papers: The parapsychological association 47th Annual Convention*, 437-439.

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.

Persinger, M.A., & Makarec, K. (1987). Temporal lobe signs and correlative behaviours displayed by normal populations. *Journal of General Psychology*, **114**, 179-195.

Radin, D.I. & Ferrari, D.C. (1991). Effects of consciousness on the fall of Dice: A meta-analysis. *Journal of Scientific Exploration*, **5**, 61-83.

Radin, D.I., & Nelson, R.D. (1989). Evidence of consciousness-related anomalies in random physical systems. *Foundations of Physics*, **19**, 1499-1514.

Rhine, L.E. (1970). *Mind over matter*. London: The Macmillan Company.

Riseman, T. (1980). *Understanding the I Ching*. London: HarperCollins.

Ritsema, R., & Sabbadini, S.A. (2005). *The original I Ching oracle*. London: Watkins Publishing.

Roe, C.A. (1995). Pseudopsychics & the Barnum Effect. *EJP*, **11**, 76-91.

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

Roe, C.A. (2001). *The psychology of psi: PK*. Talk given as part of the Rhine Research Center Summer Study Program, August 2001, Durham, NC.

Roe, C.A., Davey, R., & Stevens, P. (2003). Are ESP and PK aspects of a unitary phenomenon? A preliminary test of the relationship between ESP and PK. *JP*, **67**, 343-366.

Roe, C.A., Davey, R., & Stevens, P. (2006). Experimenter effects in laboratory tests of ESP and PK using a common protocol. *Journal of Scientific Exploration*, **20**, 239-253.

Roe, C.A., & Holt, N. (2005). A further consideration of the sender as a PK agent in ganzfeld ESP studies. *JP*, **69**, 113-127.

Roe, C.A., & Holt, N., (2006). The effects of strategy ('willing' versus absorption) and feedback (immediate versus delayed) on PK performance. *JP*, **70**, 49-67.

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

Rubin, L., & Honorton, C. (1971). Separating the yins from the yangs: An experiment with the I Ching. *Proceedings of presented papers: The Parapsychological Association 8th Annual Convention*, 6-7.

Rush, J.H. (1986). Findings from experimental PK research. In H.L. Edge, R.L. Morris, J. Palmer, & J.H. Rush Eds.), *Foundations of parapsychology: Exploring the boundaries of human capability* (pp. 237-275). London: Routledge & Kegan Paul.

Schmeidler, G.R. (1988). *Parapsychology and psychology*. Jefferson, NC: McFarland.

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: Scarecrow Press.

Schmidt, S., Schneider, R., Utts, J., & Walach, H. (2004). Distant intentionality and the feeling of being stared at. *British Journal of Psychology*, **95**, 235-247.

Shermer, M. (1997). *Why people believe weird things: Pseudoscience, sueperstition, and other confusions of our time*. New York: W.H. Freeman & Co.

Sherwood, S.J., & Roe, C.A. (2003a). A review of dream ESP studies conducted since the Maimonides dream ESP programme. *Journal of Consciousness Studies*, **10**, 85-109.

Sherwood, S.J., Roe, C.A., Holt, N., & Wilson, S. (2005). Interpersonal psi: Exploring the role of the experimenter and the experimental climate in a ganzfeld telepathy task. *EJP*, **20**, 150-172.

Shoup, R. (in press). Empirical pictures of time. In C.A. Roe, L. Coly, & W. Kramer (Eds) *Utrecht II: Charting the future of Parapsychology*.

Stainton Rogers, R. (1995). Q methodology. In J.A. Smith, R. Harré & L. Van Langenhove (Eds.) *Rethinking methods in psychology*. 178-192. London: Sage.

Stanford, R. (1978). Toward reinterpreting psi events. *JASPR*, **72**, 197-214.

Storm, L., & Thalbourne, M.A. (1998-99). The transliminal connection between paranormal effects and personality in an experiment with the I Ching. *EJP*, **14**, 100-124.

Storm, L., & Thalbourne, M.A. (2001). Studies of the I Ching: I. A replication. *JP*, **65**, 105-124.

Thalbourne, M.A. (1994). Postscript to an experiment with the I Ching: How many changing lines? *EJP*, **10**, 130-135.

Thalbourne, M.A., Delin, P.S., Barlow, J.A., & Steen, D.M. (1992-3). A further attempt to separate the Yins from the Yangs: A replication of the Rubin-Honorton experiment with the I Ching. *EJP*, **9**, 12-23.

Wilhelm, R. (1951/2003). *I Ching or book of changes*. London: Penguin Books.

Yang, W., & Sandifer, J. (2003). *The authentic I Ching*. London: Watkins Publishing.