ISSN: 1445-2308

Australian Journal of Parapsychology
Volume 9, Number 1, pp. 111-142

Investigations of the *I Ching*: II. Reliability and Validity Studies

BY LANCE STORM

Abstract: In six studies on the *I Ching*—an ancient Chinese system of divination—successful predictions of first- and second-hexagrams (i.e., hexagram hitting) based on pre-selections of corresponding descriptor-pairs have ranged from chance, to significantly above chance. No significant effect below chance has ever been found. Hexagram hitting has been predicted by measures such as paranormal belief, time perspective, and meaningfulness. Storm (2008a) found a near-significant aggregate hexagram hit rate of 27%. Though these results are encouraging, there has been no assessment of the reliability and validity of the main test instrument used in the *I Ching* studies, the Hexagram Descriptor Form (HDF). To test the validity of the HDF, three control methods were tested against the experimental method. Taking first- and second-hexagram hit rates together, three out of 22 tests on the experimental method (14%) were significant or near-significant. Three significant or near-significant outcomes out of 66 control tests (4.5%) were attributed to chance. Inter-rater reliability was tested using two *I Ching* experts who judged the 64 descriptor-pairs of the HDF for suitability against their corresponding hexagram readings. The correlation between judges' ratings was not significant (the mean rating ranged between 60% and 82%). Using the pooled data of six studies, the HDF was tested for possible selection and outcome biases. A selection bias was found, but no outcome biases were found. The *I Ching* and the HDF were considered suitable for parapsychological research.

Keywords: ESP, PK, *I Ching*, meaningfulness psi, reliability, synchronicity, validity.

INTRODUCTION

The *I Ching* is an ancient Chinese system of divination. *I Ching* users generate a hexagram (a six-line symbol) from a pool of 64 hexagrams by

throwing three coins, six times (one throw for each line). Each hexagram has a unique reading associated with it. Readings are meant to answer a pressing question or provide forecasts (i.e., advice for the future). The *I Ching* has been investigated a number of times to determine an ostensible paranormal influence in its process (e.g., Rubin & Honorton, 1971, 1972; Thalbourne, 1994; Thalbourne, Delin, Barlow, & Steen, 1992-1993), with mixed results (see Storm & Thalbourne, 2001a, for a review).

In their initial study, Storm and Thalbourne (1998-1999) hypothesised that hexagram outcomes could be predicted in advance. The number of predicted hexagram outcomes in a sample (expressed as a proportion correct) would need to be above chance.[1] Participants would pre-select 16 descriptor-pairs (i.e., ¼ of 64) that correspond to hexagram readings, but selections had to be based on feeling states rather than on mere whim as is oftentimes the case in number-calling tasks (e.g., dice-throwing). However, the same underlying assumption in dice-throwing experiments was made in the *I Ching* studies—namely, that a similar kind of anomalous process might operate while using the *I Ching* (i.e., ESP, PK, or both).

There are two types of hitting—first-hexagram hitting (any of 64 possible outcomes) and second-hexagram hitting, where the second-hexagram is derived from the 'changing line(s)' of the first hexagram (second hexagrams can only be one of the 63 remaining hexagrams).[2] Significant hexagram hit rates were found in two studies by Storm and Thalbourne (1998-1999, 2001a). No significant psi-missing effect has ever been found (Storm and Thalbourne, 2001b; Storm, 2002, 2003, 2006, 2008a; Thalbourne & Storm, 2002-2005).

However, it is yet to be shown conclusively that hexagram outcomes can be determined in advance because hit-rates have fluctuated on both sides of the chance baseline, ranging from 22% to 35% for first-hexagram hitting across six studies, and 22% to 33% for second-hexagram hitting across the same six studies. And although the aggregate hexagram hit-rate across six studies (see Storm, 2008a), where first- and second-hexagram hitting is combined, is a near-significant 27%, thus suggesting an anomalous effect, the four studies out of six that have produced hit-rates above chance were all nonsignificant.

As these results suggest, the *I Ching* effects have tended to be weak, and there has been some speculation as to how the 'slight' effect might be

---

[1] The binomial (exact) test is used to calculate the proportion correct (i.e., the hit rate) and the *p* value (a hit is designated "1" and a miss is designated "0").

[2] Coin throws of three-of-a-kind generate changing lines. A changing line changes a line in the first hexagram from "broken" to "unbroken" or vice versa, thus resulting in a second hexagram (any or all of the six lines in the first hexagram can be changing lines).

strengthened (Storm, 2008b), but it is clear that the first port of call in any serious criticism and investigation of alleged psi effects, especially when they are weak, is the methods and materials used to elicit those effects. As it stands, no reliability or validity tests have been conducted on the primary instrument used in all *I Ching* studies, the so-called Hexagram Descriptor Form (HDF; see APPENDIX A). The present study is the second of a two-part study (see Storm, 2008a, for Part 1) aimed at rectifying that situation.

RELIABILITY AND VALIDITY ANALYSES

Until recently, due to the previously limited size of the *I Ching* cumulative database, it has not been possible to conduct thorough tests on the reliability of the HDF. There are now 793 cases in the *I Ching* database, pooled from six studies, so that such tests can be conducted. Testing would include analyses of inter-rater reliability, and response biases that might manifest due to preferences participants might have based on positive and negative assumptions about the adjectives that compile the set of HDF descriptor-pairs.

Since the validity of an instrument cannot exceed its reliability, it was also deemed necessary to consider validity issues in this study. Procedures for testing the HDF are outlined in more detail in the METHODS section, but a number of issues to do with *I Ching* methodology first need clarification.

The I Ching *Coin-Throwing Methodology—Testing the Methods*

There are two commonly used *I Ching* methods that produce hexagrams from throws of three coins. Only one of them (see Wing, 1979, 1982) has been used in all six *I Ching* studies, where the outcomes of the coin throws are interpreted thus (where H = Head; and T = Tail):

HHT =  **(unbroken line);**

HTT =  **(broken line);**

HHH =  **(broken changing line);**

TTT = **(unbroken changing line).**

However, Hazel (1990) reverse scores the coin outcomes; thus:

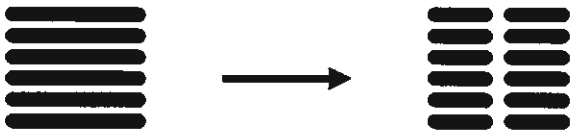
HTT = ▬▬ **(unbroken line);**

HHT = ▬ ▬ **(broken line);**

TTT = ▬ ▬ ● **(broken changing line);**

HHH = ▬▬ ● **(unbroken changing line).**

The two methods produce different hexagram outcomes. For example, Hexagram 1 using Wing's (1982) method becomes Hexagram 2 using Hazel's (1990) method (see Figure 1). This dichotomy could create confusion for the novice *I Ching* user. Hazel (1990) claims his method works best for him, and he adds that it is up to the individual to "decide" on a method and "stick" to it or "readings will be inconsistent and unreliable" (p. 6).



#1: Chi'en/ Creative *#2: K'un/ Receptive*

Figure 1. Hexagram 1 (*Chi'en: Creative*) using Wing's coin-throwing method becomes Hexagram 2 (*K'un: Receptive*) using Hazel's coin-throwing method.

There would be no dispute that coin-throwing in a given method can be relied upon to produce the same hexagram on every trial if each of the outcomes of the six coin throws are the same on every trial, and each is interpreted in the same way in every trial. However, it may be too much to claim that one method, in and of itself, is more likely than another to elicit a psi effect (and we are hypothesising that psi is involved in the *I Ching* process) because, parapsychologically, that is an issue that is primarily going to depend on the user. Thus, if we are agreed that two procedurally similar methods are equal under the sun as far as psi and chance are concerned, then it may also be too much to claim that a reading can be more pertinent by one method than another for the same set of coin outcomes (see again, Figure 1). Thus, a question mark hangs over the validity of the *I Ching* system for two reasons: (i) it is a moot point whether psi *is* the operative factor in the *I Ching* process, and (ii) the user *can* produce

different hexagrams (and therefore different readings) for the same question according to the method.[3]

Hazel's advice may speak in part to the issue of reliability, but not necessarily the issue of validity, as just argued. Therefore, the empiricist would be expected to test the various methods to see if there is objective evidence that a given *I Ching* method is not only reliable for the user *as a matter of choice*, as Hazel states, but is also valid on the basis that it elicits *purposeful* or *functional* outcomes, which we would argue can be determined statistically.

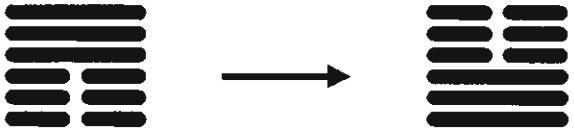
As an aside, it is also a moot point whether the validity of divinatory systems are wholly investigable, especially if psi is involved. To date, paranormal processes are unexplained and defy current scientific theories and rational understanding. More generally, the dispute over the validity of psi is still unresolved, let alone the methods of proving it (Alcock, 1981, 1998). Proof, however, lies in the evidence. If *I Ching* 'hitters' (those successful at predicting a hexagram), blind to the fact that they got a 'hit', were to rate their hexagram readings as more *meaningful* than 'missers' (those unsuccessful at predicting their hexagram, blind to the fact that they 'missed'), one might conclude that something other than a statistical anomaly pertains for hitters since it would seem that something purposeful or functional is effected for them. This meaningfulness effect was tested by Storm (2008a). The mean meaningfulness rating was significantly higher for hexagram hitters (73%) than hexagram missers (65%). Storm also found no evidence of a sheep-goat/hitting interaction effect that might explain the disparity. The effect was considered an example of synchronicity (meaningful coincidence), which Jung (1960, 1989) proposed was involved in the *I Ching* process. It might therefore be said, at least in Storm's study, that some degree of validity was demonstrated insofar as benefit to the user was determined objectively.

Since, however, there is yet no clear-cut evidence that psi operates in the *I Ching* process, it would be wise to conduct a series of control tests on the two methods just described, with Hazel's method acting as a control to Wing's method, the latter of which has been the experimental method of choice in all six *I Ching* studies to date involving the author. Such testing might reveal something more about the previously hypothesized *I Ching* effects. Relevant hypotheses will be given shortly, but continuing the line of argument, tradition holds that the coin outcomes must be recorded 'bottom-up'. The ancient Chinese argued that building a hexagram is similar to "build[ing] . . . a house" (Brennan, 2000, p. 51): One must start with a

---

[3] Besides the two two-coin methods, there are a number of four-coin methods, and even a six-coin method (see Lei-Li, 2001).

foundation, then erect the walls, and then put on a roof. Notwithstanding the fact that one might be violating the traditional approach, the coin outcomes could just as easily be recorded 'top-down' (see Figure 2).



#12: P'i/Standstill *#11: T'ai/Peace*

Figure 2. Hexagram 12 (*P'i: Standstill*), whether Wing's or Hazel's method is used (i.e., coin-outcomes are recorded 'bottom-up'), becomes Hexagram 11 (*T'ai: Peace*) if the same coin outcomes are recorded 'top-down'.

Testing these top-down pseudo-methods as control methods (i.e., mirroring the hexagrams through a vertical flip) would be an acid test of Hazel's assumption that a method, once adopted and accepted, supersedes all other methodological considerations. The following hypotheses are therefore proposed:

1. **First-hexagram hitting:** (i) using Wing's method, hitting is at a rate *greater than* MCE (where $P_{MCE} = .250$). Hitting is *at chance*, (ii) using Hazel's method; (iii) using Wing's method, and then mirroring the hexagram (i.e., recording coin outcomes top-down), and (iv) using Hazel's method, and then mirroring the hexagram.

2. **Second-hexagram hitting:** (i) using Wing's method, hitting is at a rate *greater than* MCE for first-hexagram 'hitters' ($P_{MCE} = 15/63 = .238$) and first-hexagram 'missers' ($P_{MCE} = 16/63 = .254$), Hitting is *at chance* for first-hexagram 'hitters' and first-hexagram 'missers', (ii) using Hazel's method; (iii) using Wing's method and then mirroring the hexagram, and (iv) using Hazel's method, and then mirroring the hexagram.

These two hypotheses are tested on the *I Ching* cumulative database ($N = 793$). Note that in all six studies, participants had been instructed to use Wing's method and record coin throws bottom-up. Participants had not been informed of other methods. The only factors that are changed for the present study are (i) the interpretation of coin throws, and (ii) the configuration of the lines (i.e., bottom-up vs. top-down).

If the control tests produce non-significant results, whereas hit rates are significant using Wing's method—as tested in Hypotheses 1(i) and 2(i)—then we might claim that the experimental method is a valid means of producing *I Ching* effects, but only because participants have knowingly accepted Wing's method *holus bolus*. The assumption is made that psi is more likely to manifest if participants are committed to the process, which should be the case in only one method (i.e., Wing's). Such test results, however, would not invalidate other coin-throwing methods to which users might commit. Nor would the results necessarily reflect, in whole or in part, on how the *I Ching* functions as a divinatory system.

Judging the Descriptor-Pairs of the I Ching *Hexagram Descriptor Form (HDF)*

One way of evaluating the HDF in terms of its content validity is to test the inter-rater reliability of the descriptor-pairs on the assumption that high concordance between judges is the result of good matching between descriptor-pairs and readings. Two expert users of the *I Ching* (i.e., judges), both familiar with the *I Ching* procedure and the readings associated with the 64 hexagrams, assisted in this assessment of the HDF. The two judges were asked to review the descriptor-pairs, and give a descriptor-pair rating based on whether each descriptor-pair is representative of its associated reading. The judging process requires the two-word descriptors of each of 64 readings to be evaluated qualitatively and rated quantitatively on a visual analogue scale for degree of correspondence with their associated readings (see METHOD for details). Rater correlation and rater bias would then be assessed. The following hypotheses are proposed:

3. There is a positive relationship between the two independent judges' ratings of the 64 descriptor-pairs.

4. There is no difference between paired judges' ratings of the 64 descriptor-pairs.

Testing Selection and Outcome Biases in the HDF

Due to individual differences, participants will generally express idiosyncratic, personal preferences when selecting descriptor-pairs on the HDF. However, some preferences in a given sample may constellate around specific descriptor-pairs, more so than others, thus manifesting as a bias in the distribution of descriptor-pair selections. Such a bias would suggest that

a limited number of psychological responses prevail over others as a result of similarities, rather than differences, among participants.

It is also possible, through paranormal influence, that the distribution of hexagram outcomes may show a preferential outcome. Such a bias may indicate trends for need-fulfillment or knowledge-seeking, which may only be satisfied in the participant by paranormally generating specific hexagrams to garner specific *I Ching* readings. If such an effect is possible, it can be indicated as either hit rates above MCE or as a selection bias where some hexagrams are generated more than others.

Generally, selection and outcome biases have not been considered in past *I Ching* studies, but Storm and Thalbourne (1998-1999) did consider response rates on two exemplary descriptor-pairs to ascertain whether or not at least some descriptor-pairs might solicit selection biases. It was thought that Hexagram #12 ('Stagnant, Unassisted') might be considered too negative in tone to be selected by participants, whereas Hexagram #14 ('Supreme, Successful') might be considered 'positive', thus encouraging a selection bias. However, it was found that the former was selected *more* often than the latter—a counter-intuitive finding. The statistical difference, though, between these two outcomes could not be tested due to insufficient sample size (N = 93). As Thalbourne et al. (1992-1993) more generally argued, it is only from an analysis of all hexagram choices that an overall bias could be discerned, although that would require a sample "of at least 300 readings . . . before it was really feasible to look at the distribution as a whole" (p. 20). As far as outcome biases are concerned, none of the *I Ching* studies cited above had considered the distribution of outcome hexagrams. Thalbourne (1998-1999) did test his personal cumulative record of 696 hexagrams, and found no significant deviation in hexagram outcomes, but research in this area is minimal to say the least. Therefore, the following hypotheses are proposed:

5. The frequency distribution of descriptor-pair selections is not flat.

6. The frequency distribution of hexagram outcomes is not flat.

METHOD

Judges

Two judges were recruited, both male and middle-aged. They are frequent expert users of the *I Ching*. Judge One lectures on the *I Ching*, and is a member of the South Australian Jung Society; Judge Two has spent over 20 years studying the *I Ching*. The judges are unknown to each other, and their anonymity was maintained throughout the study.

Materials

Judges were each sent a booklet containing a set of hexagram readings drawn from Wilhelm (1989) and Hazel (1990). Each hexagram had both readings on one page. At the bottom of each page was a 10-centimeter visual analogue scale (VAS) ranging from 0% to 100%. In the front of the booklet were instructions to the judges (see APPENDIX B). Two measures were also included: the Social Desirability Scale (SDS; Crowne & Marlowe, 1960) and the Uncritical Inference Test (UIT; Haney, 1954). The SDS is a measure of the degree to which a participant only responds in a way he/she deems is desired by society, and the UIT measures the degree to which the test participant makes unwarranted conclusions.

Procedure

Evaluation and inter-rater reliability of the I Ching *Descriptor-Pairs.* The judging process required the two-word descriptors of each of 64 readings to be qualitatively evaluated and scored quantitatively on 64 VASs for degree of correspondence with its associated reading. (Scoring instructions for the judges are given in APPENDIX B.) Rater-reliability and rater-bias were then assessed.

Attempts at measuring judging styles were made using the SDS and the UIT. Scores on these tests might indicate biased judging due to judges (i) projecting favorable images of themselves by trying to please the experimenter L.S. (who is also the co-designer of the HDF), and/or (ii) making unconscious but incorrect or inaccurate inferences based on the information available in the *I Ching* readings.

Methods of evaluation. Raw data was analysed statistically using *SPSS* and the online *VassarStats* Exact Binomial Calculator (Lowry, 1998-2006). Specific statistical tests for all hypotheses include ANOVA, Pearson's r, independent samples t-test, and chi-square goodness-of-fit. The datasets are the six *I Ching* studies, assessed separately and combined ($N = 793$).

RESULTS

Scale Scores

Social Desirability Scale (SDS). The score for Judge One was 19, and the score for Judge Two was 14. The judges' scores were not extreme compared to a large-sample mean score of 15.08 (Andrews & Meyer, 2003). The Z score transformation formula was used to calculate judge's

standardised scores: $z = [X - M]/SD)$, where X = raw score, M = sample mean score, and SD = standard deviation. Judge One: $z_1 = [19 - 15.08]/5.50 = 0.71, p_1 = .239$; Judge Two: $z_2 = [14 - 15.08]/ 5.50 = -0.20, p_2 = .421$.

Uncritical Inference Test (UIT). The score for Judge One was 64, and the score for Judge Two was 39. The judges' scores were not significantly different from the sample mean score for males of 50.30 (SD = 10.80) in the study by Tobacyk and Milford (1982). Judge One: $z_1 = [64 - 50.30]/10.80 = 1.27, p_1 = .102$; Judge Two: $z_2 = [39 - 50.30]/10.80 = -1.05, p_2 = .147$.

Planned Analyses

Hypothesis 1: First-hexagram hit rates. The pooled data from the six studies yielded a total number of first hexagrams of 793. Table 1 lists first-hexagram hit rates for the experimental method (Wing's Method) and the three control methods. As hypothesised, the hit rate on first-hexagram hitting using Wing's Method was above MCE, but not significantly: P = 26.2% (exact p = .223).

Table 1
First-Hexagram Hit Rates For The Four Methods (Six Studies Combined)

Method	Full Sample			
	N	Hits	*P***	Exact *p*
1. Wing's Method (bottom-up)*	793	208	.262	.223
2. Hazel's Method (bottom-up)	793	206	.259	.752
3. Wing's Method (top-down)	793	203	.256	.669
4. Hazel's Method (top-down)	793	189	.238	.787

* Method 1 = experimental method (Methods 2, 3, & 4 = Controls); ** P_{MCE} = .250

Two control hit rates were above MCE, and one was below MCE, but none were significant. All three control hit rates were *lower* than the experimental hit rate, but the differences between all four hit rates were not significant using a repeated-measures ANOVA, $F(3, 2376) = 0.51, p = .678$.

Table 2 lists hit rates for each of the six studies by each of the four methods (i.e., 6 × 4 = 24 tests in total). Hit rates on Wing's Method were above chance four times out of six (67%). One study was significant, and one was marginally significant (p = .017, and p = .070). It was hypothesised that the control methods would produce null results (i.e., hit rates *at* or *below* chance), and 8 out of 18 (56%) *were* at or below MCE.

Table 2
First-Hexagram Hit Rates For The Four Methods Across Six Studies

Method	Storm & Thalbourne (1998-1999)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	93	30	.322	.070
2. Hazel's Method (bottom-up)	93	25	.269	.376
3. Wing's Method (top-down)	93	24	.258	.468
4. Hazel's Method (top-down)	93	21	.226	.741
Method	Storm & Thalbourne (2001a)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	107	37	.346	.017
2. Hazel's Method (bottom-up)	107	33	.308	.101
3. Wing's Method (top-down)	107	28	.262	.426
4. Hazel's Method (top-down)	107	30	.280	.266
Method	Storm (2002)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	43	11	.256	.523
2. Hazel's Method (bottom-up)	43	15	.349	.096
3. Wing's Method (top-down)	43	7	.163	.939
4. Hazel's Method (top-down)	43	6	.139	.974
Method	Thalbourne & Storm (2002-2005)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	200	43	.215	.891
2. Hazel's Method (bottom-up)	200	46	.230	.767
3. Wing's Method (top-down)	200	48	.240	.654
4. Hazel's Method (top-down)	200	36	.180	.993
Method	(Storm, 2006)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	200	49	.245	.592
2. Hazel's Method (bottom-up)	200	50	.250	.527
3. Wing's Method (top-down)	200	58	.290	.111
4. Hazel's Method (top-down)	200	52	.260	.398
Method	(Storm, 2008a)			
	N	Hits	P^{**}	Exact *p*
1. Wing's Method (bottom-up)*	150	38	.253	.494
2. Hazel's Method (bottom-up)	150	37	.247	.569
3. Wing's Method (top-down)	150	38	.253	.494
4. Hazel's Method (top-down)	150	44	.293	.130

* Method 1 = experimental method (Methods 2, 3, & 4 = Controls); ** P_{MCE} = .250

None of the 18 control tests produced significant hit rates above MCE. Out of 28 tests run (see Tables 1 & 2), 14 test results (50%) were in the directions hypothesised, which is exactly at chance. (See *Performance summary*, p. 123, for a more general analysis of hitting outcomes.)

Hypothesis 2: Second-hexagram hit rates. The pooled data from the six studies yielded a total number of second hexagrams of 631. Table 3 lists the two types of second-hexagram hitting for the experimental method (Wing's Method) and the three control methods. Hit rates for both types of second-hexagram hitting using Wing's Method were above MCE (P = 27.9%, p = .121; 27.0%, p = .228), but not significantly. For the six control tests, three hit rates were at or below MCE. One was significantly above chance, and this was for first-hexagram missers.

Table 3
Second-Hexagram Hit Rates For The Four Methods (Six Studies Combined)

Method	First-Hexagrams Hitters				First-Hexagrams Missers			
	N	Hits	P^{**}	Exact p	n	Hits	$P^{\dagger}$	Exact p
1. Wing's Method (bottom-up)*	172	48	.279	.121	459	124	.270	.228
2. Hazel's Method (bottom-up)	172	42	.244	.454	459	143	.311	.003
3. Wing's Method (top-down)	172	42	.244	.454	459	110	.240	.775
4. Hazel's Method (top-down)	172	32	.186	.958	459	114	.248	.627

* Method 1 = Experimental method (Methods 2, 3, & 4 = Controls); ** P_{MCE} =.238; † P_{MCE} = .254

The combined hit rate was then tested across methods (i.e., second hexagram outcomes for first-hexagram hitters were added to those of first-hexagram missers). For Wing's Method, the hit rate was above MCE (where P_{MCE} = 24.96%),[4] which approached significance, P = 27.3% (172 hits; N = 631; exact p = .099). However, the hit-rate for Hazel's Method was significantly above MCE, 29.3% (185 hits; p = .007).

Hit rates for the other two control methods were 24.1% (152 hits; p = .632); and 23.1% (146 hits; p = .815), both of which were below MCE, and neither was significant.

Using a repeated-measures ANOVA, Hazel's Method yielded a hit rate significantly higher than the other three, $F(3, 1890) = 2.86$, p = .036.

[4] $P_{MCE} = (15/63 \times 172/631) + (16/63 \times 459/631) = 24.96\%$

Table 4 lists hit rates for each of the six studies by each of the four methods for both types of second-hexagram hitting (i.e., $6 \times 4 \times 2 = 48$ tests in total). More than half of the tests on Wing's Method (i.e., 7 tests out of 12; 58%) produced hit rates above MCE, but none were significant.

More than half the control tests (i.e., 21 out of 36; 58.5%) produced hit rates at or below MCE. One hit rate was significantly above chance (p = .026), and one was marginally significant (p = .072).

Out of 60 tests run, 36 results were in the directions hypothesised (p = .08). (See the next section for a more general analysis of hitting outcomes.)

Performance summary. For significance tests on the experimental method—first- and second-hexagram hitting taken together—3 out of 22 results were significant or marginally significant (14%). Applying the 5% rule, only one of these three could be attributed to chance.

For the control methods, 3 out of 66 test results were significantly or marginally significantly above MCE (4.5%). We would expect three test results to be significant or marginally significant by chance alone.

In terms of directionality, 22 tests were run on the experimental method, and 66 were run on control methods, yielding a total of 88 tests. Of these 88 tests, 50 were in the directions hypothesised (p = .120).

Hypothesis 3: Correlation of judges' ratings. The relationship between the judges' ratings was not positive or significant, $r(63) = -.05$, p = .341 (one-tailed). When the judges' ratings were recalibrated (0 to 33 = '0'; 34 to 67 = '1'; 68 to 100 = '2'), the correlation was still not significant, $r(63) = -.02$, p = .444 (one-tailed). If the judges rated the descriptor-pairs in a mutually compatible manner, we might expect a significant and large positive correlation, but this was not the case. The hypothesis was not supported. There was no inter-rater reliability.

Hypothesis 4: Differences between judges' ratings. The mean descriptor-pair suitability rating for Judge One was 81.98% (SD = 17.02%) and for Judge Two it was 59.92% (SD = 26.33%). Thus, the HDF as a single instrument (i.e., the entire set of 64 descriptor-pairs) ranges in suitability from approximately 60% to 82%. Using the paired-samples t test, the mean difference was highly significant, $t(63) = 5.50$, $p < .001$ (two-tailed). The null hypothesis was rejected.

If, however, Judge One (the more 'generous' rater) had rated descriptor-pairs consistently higher than Judge Two, we might expect a significant correlation in Hypothesis 3 above, which was not the case.

Table 4
Second-Hexagram Hit Rates For The Four Methods Across Six Studies

Method	Storm & Thalbourne (1998-1999)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*n*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	27	9	.300	.173	52	15	.288	.333
2. Hazel's Method (bottom-up)	27	6	.222	.649	52	16	.308	.229
3. Wing's Method (top-down)	27	6	.222	.649	52	10	.192	.884
4. Hazel's Method (top-down)	27	3	.111	.972	52	10	.192	.884

Method	Storm & Thalbourne (2001a)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*N*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	27	6	.222	.649	52	15	.288	.333
2. Hazel's Method (bottom-up)	27	8	.296	.303	52	17	.327	.147
3. Wing's Method (top-down)	27	8	.296	.303	52	11	.212	.804
4. Hazel's Method (top-down)	27	8	.296	.303	52	10	.192	.884

Method	Storm (2002)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*n*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	7	2	.286	.525	28	7	.250	.591
2. Hazel's Method (bottom-up)	7	0	.000	1.000	28	7	.250	.591
3. Wing's Method (top-down)	7	1	.143	.851	28	5	.179	.875
4. Hazel's Method (top-down)	7	0	.000	1.000	28	4	.143	.950

Method	Thalbourne & Storm (2002-2005)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*n*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	36	10	.278	.347	127	30	.236	.709
2. Hazel's Method (bottom-up)	36	10	.278	.347	127	40	.315	.072
3. Wing's Method (top-down)	36	9	.250	.497	127	38	.299	.143
4. Hazel's Method (top-down)	36	5	.139	.952	127	30	.236	.709

Method	(Storm, 2006)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*n*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	41	13	.317	.157	112	35	.313	.096
2. Hazel's Method (bottom-up)	41	9	.220	.668	112	35	.313	.096
3. Wing's Method (top-down)	41	12	.293	.255	112	29	.259	.488
4. Hazel's Method (top-down)	41	8	.195	.793	112	29	.259	.488

Method	(Storm, 2008a)							
	First-Hexagrams Hitters				First-Hexagrams Missers			
	n	Hits	P^{**}	Exact *p*	*n*	Hits	$P^{\dagger}$	Exact *p*
1. Wing's Method (bottom-up)*	34	8	.235	.581	88	22	.250	.575
2. Hazel's Method (bottom-up)	34	6	.176	.853	88	28	.318	.105
3. Wing's Method (top-down)	34	6	.176	.853	88	17	.193	.928
4. Hazel's Method (top-down)	34	8	.235	.581	88	31	.352	.026

* Method 1 = experimental method (Methods 2, 3, & 4 = Controls); ** P_{MCE} =.238; † P_{MCE} = .254

Therefore, evaluation criteria appear to be different between the judges, though we may be justified in ruling out social desirability effects and uncritical inference-making since judges' scores on the SDS and the UIT were not significantly different from the relevant mean scores (see results reported above in the subsection *Scale Scores*).

Although SDS scores were not significantly different between judges,[5] it is, however, important to point out that the two judges' scores on the UIT were significantly different from each other.[6] This nonchance difference suggests two different cognitive styles which may account for the rating differential since the difference seems to be a question of abstract vs. concrete thinking. (This issue is taken up in the DISCUSSION.)

Hypothesis 5: Distribution of descriptor-pair selections. Figure 3 shows the frequency distribution of hexagrams that correspond with the descriptor-pair selections participants made on the HDF. Table C1 (see APPENDIX C) lists frequency counts for all descriptor-pairs as represented by their corresponding hexagrams. Table C1 also lists all 64 hexagrams and their ranks from least preferred (i.e., rank of 64) to most preferred (i.e., rank of 1).

All 64 descriptor-pairs were selected by participants with a minimum frequency of 56 times for the descriptor-pair corresponding to Hexagram #64 and with maximum frequency of 465 times for the descriptor-pair corresponding to Hexagram #28.

Visual inspection of Figure 3 suggests a bias in selections. The chi-square test indicates the distribution was not flat, $\chi^2(63, N = 12688) = 3303.84, p < .001$, two-sided. The hypothesis was supported.

Hypothesis 6: Distributions of hexagram outcomes. Figure 4 shows the distribution of first-hexagram outcomes. Table C2 (see APPENDIX C) lists all 64 first-hexagram outcomes in order of frequency from lowest to highest. All 64 hexagrams were generated as first hexagrams (min. = 5; max. = 22).

The chi-square test indicates the distribution was flat, $\chi^2(63, N = 793) = 48.04, p = .919$, two-sided. The hypothesis was not supported. The result shows that psi was not acting to bias the outcomes.

---

[5] Judges' scores were tested using Rosenthal and Rubin (1979) $Z_{diff.}$ formula: $Z_{diff.} = [Z_1 - Z_2]/\sqrt{2}$, where Z_1 is the z score for Judge One and Z_2 is the is the z score for Judge Two. SDS: $[0.71 - (-0.20)]/\sqrt{2} = .91/1.41 = 0.64$ ($p = .26$). Scores were not significantly different from each other.

[6] UIT: $[1.27 - (-1.05)]/\sqrt{2} = 2.32/1.41 = 1.64$ ($p = .05$). UIT scores are significantly different.

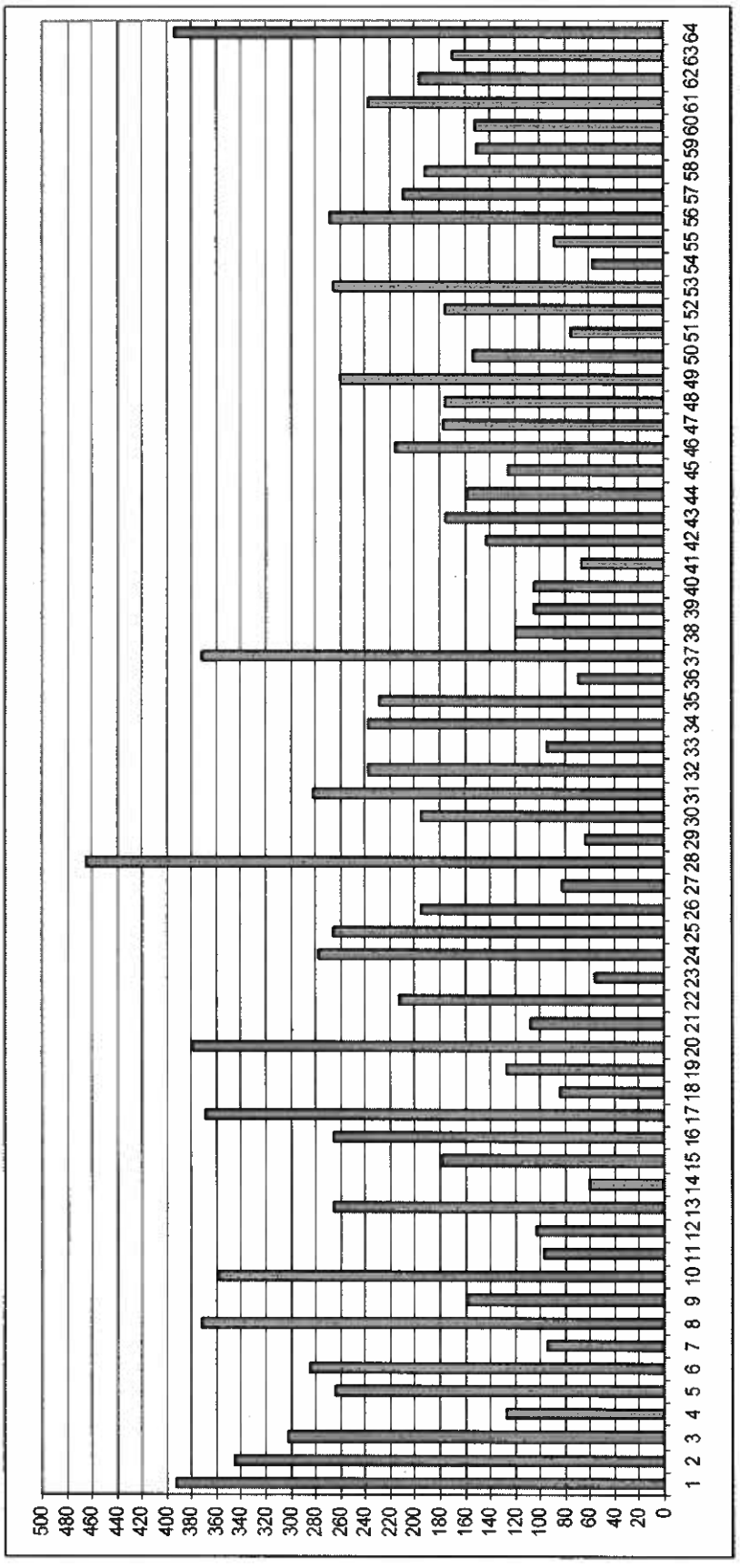


Figure 3. Distribution of hexagram selections for six pooled studies (N = 793). Least frequent was Hexagram #23 (56 cases), and most frequent was Hexagram #28 (465 cases).

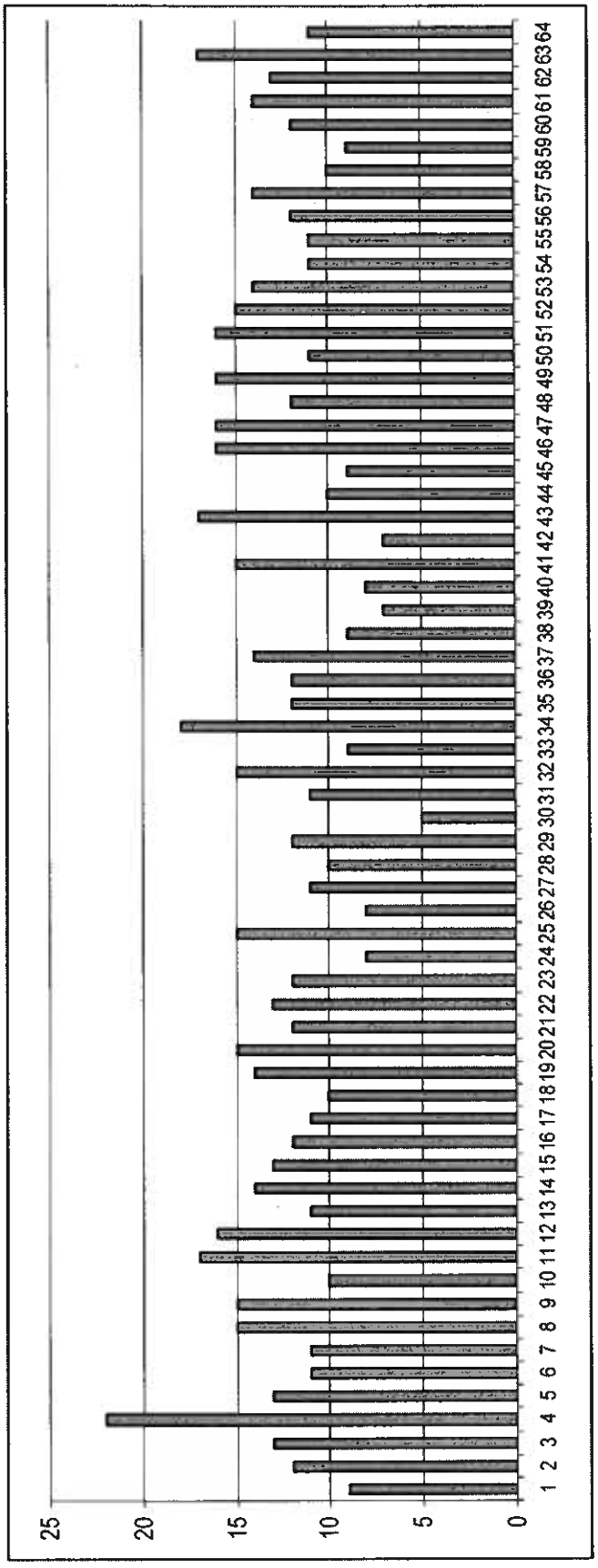


Figure 4. Distribution of first-hexagram outcomes for six pooled studies (N = 793). Least frequent was Hexagram #30 (5 cases), and most frequent was Hexagram #4 (22 cases).

Figure 5 shows the distribution of second-hexagram outcomes. Table C3 (see APPENDIX C) lists all 64 first-hexagram outcomes in order of frequency from lowest to highest. All 64 hexagrams were generated as second hexagrams (min. = 3; max. = 20).[7]

The chi-square test indicates the distribution was flat, $\chi^2(63, N = 631) = 66.51$, $p = .357$, two-sided. The hypothesis was not supported. Again, the result shows that psi was not acting in a biased way.

Two Spearman's *rho* tests were conducted to see if hit-ratios (number of hits/descriptor-pair selection counts) were related to hexagram ranks on the premise that selection biases found in Hypothesis 5 may have encouraged a paranormal outcome bias. The correlations were not significant: First hexagrams, $r_s(62) = 0.01$, $p = .459$ (one-tailed); second hexagrams, $r_s(62) = 0.03$, $p = .412$ (one-tailed). There was no psi bias.

Cumulative Record [8]

The cumulative record for hexagram hit rates across six studies is shown in Table 5 (p. 130). For the pooled databases, the overall hit rates for the two kinds of hitting (i.e., first- and second-hexagram hitting) range from 26% to 28%.

The difference between first- and second-hexagram hitting was not significant, $t(1422) = 0.44$, $p = .663$ (two-tailed). The weighted aggregate hit rates (first- and second-hexagram hitting combined) are given in Table 6.

The overall hit rate for all six-studies combined ($N = 1424$; Hits = 380) was a near-significant 27% ($p = .072$, where $P_{MCE} = 24.98\%$). [9]

---

[7] Note that all chi-square tests meet the two-part assumption for large tables: (i) *at least* 80% of cells have $k \geq 5$, and (ii) no cells have a zero count (Hinders, 2007, p. 269).

[8] Most of the information in this section are updates to Tables 5 and 6 in Storm (2008, pp. 119-120).

[9] Weighted aggregate hexagram hit rate is based on proportions of the sample, where:

$$P_{MCE} = (16/64 \times 793/1424) + (15/63 \times 172/1424) + (16/63 \times 459/1424) = 24.98\%$$

Note that the P_{MCE} value for aggregate hexagram hitting given in previous publications (see Storm, 2008a; Thalbourne & Storm, 2002-2005) were calculated using the unweighted formula:

$$P_{MCE} = [(16/64) + (15/63) + (16/63)]/3 = [.250 + .238 + .254]/3 = 24.76\%.$$

Note for both formulae that P_{MCE} falls slightly short of 25% because the two P_{MCE} values for second hexagram hitting (i.e., .238 & .254) are not equidistant from .250 (see Hypothesis 2, p. 116).

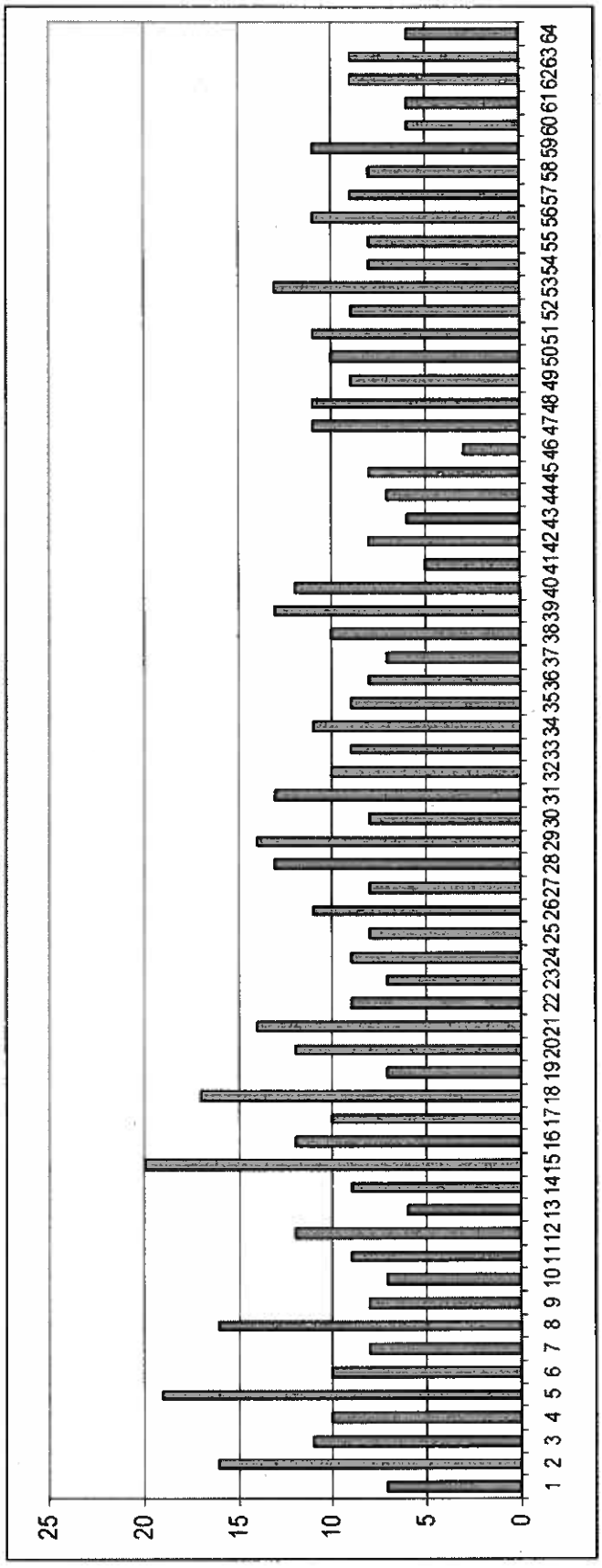


Figure 5. Distribution of second-hexagram outcomes for six pooled studies ($N = 631$). Least frequent was Hexagram #46 (3 cases), and most frequent was Hexagram #15 (20 cases).

Table 5
Hexagram Hit Rates for Six *I Ching* Studies ($N = 793$)

Study	1st Hexagrams				2nd Hexagrams (1st hexagram hitters)				2nd Hexagrams (1st hexagram missers)			
	N	Hits	% [a]	*p*	*n*	Hits	% [b]	*p*	*n*	Hits	% [c]	*p*
1. Storm & Thalbourne (1998-1999)	93	30	32	.070	27	9	33	.173	52	15	29	.333
2. Storm & Thalbourne (2001a)	107	37	35	.017	27	6	22	.649	52	15	29	.333
3. Storm (2002)	43	11	26	.523	7	2	29	.525	28	7	25	.383
4. Thalbourne & Storm (2002-2005)	200	43	22	.891	36	10	28	.347	127	30	24	.709
5. Storm (2006)[d]	200	49	25	.592	41	13	32	.157	112	35	31	.096
6. Storm (2008)	150	38	25	.494	34	8	24	.581	88	22	25	.575
Totals	793	208	26	.223	172	48	28	.121	459	124	27	.228

Note: p values are exact: [a] $P_{MCE} = .250$; [b] $P_{MCE} = .238$; [c] $P_{MCE} = .254$; [d] corrected hit rates (see Storm, 2008, p. 119).

Table 6
Aggregate Hexagram Hit Rates for the Six *I Ching* Studies ($N = 1424$)

Study	*N*	Hits	%*	*p*
1. Storm & Thalbourne (1998-1999)	172	54	31	.028
2. Storm & Thalbourne (2001a)	186	58	31	.027
3. Storm (2002)	78	20	26	.467
4. Thalbourne & Storm (2002-2005)	363	83	23	.808
5. Storm (2006)	353	97	27	.126
6. Storm (2008a)	272	68	25	.478
Totals	1424	380	27	.072

* See Footnote 9

Table 7 lists the aggregate hit rates for the control methods. As can be seen, only one control method of three produced a significant hit rate—Hazel's method (bottom-up). This significant effect is largely attributable to second-hexagram hitting, whereas Wing's method is more consistent, as the results have shown (see *Performance summary*, p. 123).

Table 7
Aggregate Hexagram Hit Rates for the Control Methods ($N = 1424$)

Study	Hits	%*	*p*
Hazel's Method (bottom-up)	391	27	.017
Wing's Method (top-down)	355	25	.528
Hazel's Method (top-down)	335	24	.903

* See Footnote 9

DISCUSSION

I Ching *Effects*

This study was primarily a test of one popular *I Ching* coin-throwing method (i.e., Wing's Method), which has been used in six studies to date. This method was the so-called *experimental* procedure for the purposes of the present study, and it was tested against another popular method (i.e., Hazel's Method) and two control methods (Wing's and Hazels' pseudo-methods where coin throws were recorded top-down). Needless to say,

Hazel's Method also served as an effective control. Thus there were three controls.

First-hexagram hitting. Using the experimental procedure (i.e., Wing's Method) to generate first hexagrams, and testing only the cumulative database (N = 793), first-hexagram hitting was above MCE at 26% (not significant). The three control methods produced no significant hit rates, but all three hit rates *were* lower than 26%, and below MCE (not significantly).

Again for the experimental method, when the six studies were considered individually, hit rates were above chance four times out of six (67%; once significantly, and once marginally significantly). However, no significant effects were found using the three control methods (6 studies × 3 control methods = 18 tests, all of which were not significant).

Second-hexagram hitting. From the cumulative record (N = 631), tests on Wing's Method produced hit rates ranging from 27% for first-hexagram missers to 28% for first-hexagram hitters. Neither hit rate was significant, but the combined hit rate did approach significance: 27% (p = .099).

For the experimental method, when the six studies were considered individually, hit rates were above chance seven times out of twelve (58%), though none were significant.

The significant above-chance hit rate of 31% (p = .007) using Hazel's Method is likely to be a chance result (see next section). Also, this hit rate of 31% was the hit rate for first-hexagram missers, whereas for first-hexagram hitters, the hit rate was a rather low and non-significant 24%. At least Wing's Method was consistent, having produced above-chance hit rates for *both* types of second-hexagram hitting as just indicated (i.e., 27% and 28%).

Finally, aggregate hitting (first- and second-hexagram hitting combined) was at or above MCE five times out of six (see Table 6).

Summary. For the experimental method, on first- and second-hexagram hitting, there were three significant or marginally significant hit rates out of 22 tests (14%). Applying the 5% rule, one of these three could be attributed to chance. However, out of 66 control tests, only three results (4.5%) were significant or marginally significant which, theoretically, could all be attributed to chance.

In summing up the four 'methods' (actually 1 method + 3 controls), any consistent coin-throwing method would, by definition, be reliable, but it does seem that the statistical evidence has helped validate only Wing's Method as the more consistent method, even when compared to test results for Hazel's Method.

Can it be said, however, that the *I Ching* (or any divinatory system) is validated once psi as a statistical anomaly is elicited? We can only answer that question in the affirmative if we can also provide statistical evidence of intentionality and meaningfulness, both of which must also be crucial factors in a divinatory system. Intention is evident in the participant's belief in, and attitudes toward the method and the outcome, as well as the question that is posed, and the relevant mental and emotional states which may fuel that question. The parapsychologist cannot deny that paranormal influences on the system may result from intentionality (see also, Jung, 1960, 1989, for a similar view). Intentionality—a form of belief—has been demonstrated in the sheep-goat effect in a number of studies featuring the *I Ching* studies (see Storm & Thalbourne, 1998-1999, 2001a,b; Thalbourne & Storm, 2002-2005, Storm, 2002, 2003, 2006, 2008a).

As for meaningfulness, it is primarily a subjective evaluation on the part of the participant, dependent upon whether the reading makes sense. But meaningfulness can be given some degree of objectification through statistical analysis. Storm's (2008a) previous finding of a meaningfulness effect emphasises the likelihood that the *I Ching* can be validated if purposeful or functional psi effects are shown to underlie, and emerge from, the process.

The I Ching *Hexagram Descriptor Form (HDF)*

The judges. The judges' appeared to use different judging criteria (see results for Hypotheses 3 & 4). The significant difference on the two judges' UIT scores shows how easy it can be to find extreme cognitive styles just by chance alone. The rating differential appears to be the result of abstract vs. concrete thinking, so that one judge might have been overly-critical of descriptor-pairs while the other was necessarily flexible, or one over-interpreted while the other was necessarily specific. Such differences may mean the inter-rater reliability exercise was of dubious merit. But if it is fair to ignore these UIT findings, and this author doubts it, we must concede that the HDF might be unreliable. In future *I Ching* studies it would be wise to include an alternative HDF compiled by *I Ching* experts, but any new HDF on offer must also be subjected to tests on inter-rater reliability.

Judges' descriptor-pair ratings and hexagram outcomes. The mean suitability ratings for the HDF descriptor-pairs were 60% and 82% depending on the judge. These two mean ratings may suggest many things to many people. It may be difficult to answer the question, Is the HDF suitable for parapsychological investigations of the *I Ching*? Clearly, the HDF could never be aligned with the traditional purposes of the *I Ching*,

but that was never the aim of the experimenters. In the experimental situation, the HDF was only ever a means to an end, and to some degree, that end has been accomplished. Psi effects, often referred to as "*I Ching* effects" in studies dating from 2002 (see Thalbourne & Storm, 2002-2005), were demonstrated in a number of different ways that could only be achieved by using the HDF. Specifically, the statistical evidence seems to suggest that a sufficient number of participants can generate hexagram readings that meaningfully parallel their personal preferences (see Storm 2008a). These preferences are based on selections of descriptor-pairs in response to the prompt "Lately, or right now, I feel . . .," which is thus a key component of the HDF. Indirectly, the HDF played a mediating role in validating participants' psi beliefs. Participants classed as sheep, tended to be first-hexagram hitters, which means they tended to select descriptor-pairs that matched their outcome hexagrams more often than goats (see again, Storm, 2008a).

Notwithstanding these effects, it is emphasised that the shortfall of at least 18% (possibly as high as 40%) implies some degree of ambiguity in the HDF such that misdirection effects might have taken place as a result of inappropriately labelled descriptor-pairs. If there are some number of descriptor-pairs that do mislead or misdirect due to inappropriate labelling, we cannot be sure whether they have an effect on participants' psi, but there is no evidence to suggest that they do. Even though the distribution of descriptor-pairs was not flat (i.e., there was a selection bias—see Figure 3), that finding only indicates normal psychological biases in participants (see the next section), not paranormal biases, and the two flat distributions for first- and second-hexagram outcomes (see Figures 4 & 5) also suggest no psi-conducive preferential effects.

Descriptor-pair selections as preferences vs. non-preferences. Identifying the psychological sources of the selection bias just mentioned may require not only further quantitative analyses, but also qualitative analyses. We can say, based on participants' preferences (and non-preferences)[10] that the typical participant is most likely to feel stressed, challenged (Hexagram #28), hopeful, reserved (Hexagram #64), and creative, motivated (Hexagram #1), and *least* likely to feel crest-fallen, disabled (Hexagram #23), subordinate, disadvantaged (Hexagram #54), and supreme, successful (Hexagram #14). It is noted that Hexagram #14 was the least preferred as early as Storm and Thalbourne's (1998-1999) initial *I Ching* study. Perhaps

[10] Five percent of the most preferred descriptor-pairs, and 5% of the least preferred descriptor-pairs were taken for this analysis. That is, three descriptor-pairs from both the 'top' and the 'bottom' of the list (see Table C1, Appendix C).

participants are too modest to admit to feelings of supremacy and/or success, besides which, participants would be aware that these states are relative.

While it seems that a majority of participants felt free to admit they were stressed and/or challenged, it was a minority who either did not feel crest-fallen and/or disabled, or were not prepared, or able, to make such a judgement. These kinds of biases can be expected. Stressed or challenged individuals would seek advice, whether they consciously know it or not, and they might solicit advice from any source on offer, even those which conventional society might deem questionable or unproven such as the *I Ching*. Likewise, many would agree that it is human nature not to admit to failure or disability, even to oneself. Similar psychological arguments can be put forward for the other *most* preferred and *least* preferred descriptor-pairs.

Conclusion

It is acknowledged that parametric testing, insofar as this can be done on parapsychological measures, should be undertaken *before* running the actual study so that the experimenter may feel confident in the reliability and validity of his/her test instruments. However, in terms of reliability, it should be noted that from the outset the HDF was designed with considerable care for use in the author's Honours thesis (see Storm & Thalbourne, 1998-1999). As for the *I Ching* procedure used in all six studies, there is no doubt that the coin-throwing method is consistent and reliable.

As far as validity is concerned, the *I Ching*, as a divinatory system, may be difficult to assess. Nevertheless, it may be the case that manifested psi or synchronicity during the *I Ching* process helps clarify the issue of validity. In other words, we might be able to draw an equivalence between psi and divination, but given the nature of psi, we may only be substituting one mystery (i.e., divination) with another (i.e., psi), thus undermining our argument.

In further defense of the seemingly belated tests on the HDF in the present study, a number of these tests could only be conducted once sufficient data had been accumulated. Tests on this substantial database seemed to indicate a statistical anomaly, regardless of whether some degree of ambiguity hangs over the instrument used to produce it—namely, the HDF (specifically referring here to the issue of descriptor-pair suitability based on our two judges' data). Primarily, the cumulative record indicates some relatively consistent above-chance hexagram hit rates which, by and large, were only generated by the method in question—namely, Wing's

Method. Therefore, it is argued, that at least one *I Ching* method (i.e., Wing's) and the HDF are suitable for parapsychological research insofar as neither inhibits psi, but both, in fact, seem to provide a mechanism that may facilitate the psi process. It is duly noted, however, that refinement of the HDF may lead to even greater yields of significant results.

ACKNOWLEDGEMENTS

Preparation of this article was supported by a grant from the Bial Foundation.

The author thanks Dr. Michael A. Thalbourne and Professor Suitbert Ertel for helpful comments and advice.

REFERENCES

Alcock, J. E. (1981). *Parapsychology, science or magic?* Oxford: Pergamon Press.

Alcock, *J. E. (1998). Science, pseudoscience, and anomaly.* Behavioral and Brain Sciences, *21*, 301-310.

Andrews, P, & Meyer, R. G. (2003). Marlowe-Crowne Social Desirability Scale and short Form C: Forensic norms. *Journal of Clinical Psychology*, *59*(4), 483-492.

Brennan, J. H. (2000). *The magical I Ching.* St. Paul, MN: Llewellyn Publications.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Haney, W. (1954). Measurement of the ability to discriminate between inferential and descriptive statements. (Doctoral dissertation, Northwestern University). *Dissertation Abstracts International*, *14*, 405-406. (University Microfilms, No. 7037)

Hazel, R. (1990). *Consulting the coins: A new age interpretation of the I Ching.* Port Melbourne: Lothian.

Hinders, D. C. (2007). *5 Steps to a 5 AP Statistics 2008-2009.* New York: McGraw-Hill Professional.

Jung, C. G. (1960). *The structure and dynamics of the psyche.* Princeton, NJ: Princeton University Press.

Jung, C. G. (1989). *Foreword to The I Ching or book of changes.* In R. Wilhelm, *The I Ching or book of changes* (pp. *xxi-xxxix*). C. F. Baynes. (trans.). Princeton, NJ: Princeton University Press.

Horowitz, L. M., Inouye, D., & Siegelman, E. Y. (1979). On averaging judges' ratings to increase their correlation with an external criterion. *Journal of Consulting and Clinical Psychology, 47*(3), 453-458.

Lowry, R. (1998-2006). *VassarStats exact binomial calculator.* Retrieved June 17, 2006, from http://faculty.vassar.edu/lowry/binomialX.html.

Lei-Li, M. (2001). Consulting the *Yi Jing* (*I Ching*) Oracle. Retrieved April 30, 2009, from http://www.dentato.it/iching/howtocast.htm.

Rosenthal, R., & Rubin, D. B. (1979). Comparing significance levels of independent studies. *Psychological Bulletin, 86*, 1165-1168.

Rubin, D. B., & Honorton, C. (1971). Separating the yins from the yangs: An experiment with the *I Ching*. *Journal of Parapsychology, 35*, 313-314.

Rubin, D. B., & Honorton, C. (1972). Separating the yins from the yangs: An experiment with the *I Ching.* In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.), *Proceedings of the Parapsychological Association 1971, No. 8* (pp. 6-7). Durham, NC: Parapsychological Association.

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

Storm, L., & Thalbourne, M. A. (2001b). Studies of the *I Ching*: II. Additional analyses. *Journal of Parapsychology, 65*, 291-309.

Storm, L. (2002). A parapsychological investigation of the *I Ching*: Seeking psi in an ancient Chinese system of divination. *Australian Journal of Parapsychology, 2*, 44-62.

Storm, L. (2003). Research note: The pro attitude and the "sheep-goat" effect in the *I Ching* database. *Australian Journal of Parapsychology, 3*, 147-152.

Storm, L. (2006). A parapsychological investigation of the *I Ching*: The relationships between psi, intuition, and time perspective. *Journal of Parapsychology, 70*, 121-141.

Storm, L. (2008a). Investigations of the *I Ching*: I. Relationships between psi, time perspective, paranormal belief and meaningfulness. *Australian Journal of Parapsychology, 8*, 103-127.

Storm, L. (2008b). Synchronicity and psi. In L. Storm (Ed.), *Synchronicity: Multiple perspectives on meaningful coincidence* (pp. 273-296). Pari, Italy: Pari Publishing.

Storm, L., & Thalbourne, M. A. (1998-1999). The transliminal connection between personality and paranormal effects in an experiment with the *I Ching*. *European Journal of Parapsychology, 14*, 100-124.

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

Storm, L., & Thalbourne, M. A. (2001b). . Studies of the *I Ching*: II. Additional analyses. *Journal of Parapsychology, 65*, 291-309.

Thalbourne, M. A. (1994). Postscript to an experiment with the *I Ching*: How many changing lines? *European Journal of Parapsychology, 10*, 130-145.

Thalbourne, M. A. (1998-1999). Research Note: In search of psi in the *I Ching* meta-hexagram. *European Journal of Parapsychology, 14*, 125-129.

Thalbourne, M. A., Delin, P. S., Barlow, J. A., & Steen, D. M. (1992-1993). A further attempt to separate the yins from the yangs: A replication of the Rubin-Honorton experiment with the *I Ching*. *European Journal of Parapsychology, 9*, 12-23.

Thalbourne, M. A., & Storm, L. (2002-2005). A further study of psychopraxia using the *I Ching*. *International Journal of Parapsychology, 13*.

Tobacyk, A., & Milford, G. (1982). Criterion validity for Ellis' Irrational Beliefs: Dogmatism and uncritical inferences. *Journal of Clinical Psychology, 38*(3), 605-607.

Wilhelm, R. (1989). *The I Ching or book of changes*. C. F. Baynes (trans.). Princeton, NJ: Princeton University Press.

Wing, R. L. (1979). *The* I Ching *workbook*. New York: Doubleday.

Wing, R. L. (1982). *The illustrated* I Ching. Wellingborough, Northamptonshire, England: Aquarian Press.

Anomalistic Psychology Research Unit
School of Psychology
University of Adelaide
Adelaide, SA 5005
AUSTRALIA

Email: lance.storm@adelaide.edu.au

APPENDIX A

Lately, or right now I feel:

[ ] Creative, Motivated	[ ] Adaptable, Helpful	[ ] Retroactive, Concerned	[ ] Changeable, Transformed
[ ] Receptive, Accepting	[ ] Negligent, Habituated	[ ] Empowered, Tested	[ ] Spiritual, Fulfilled
[ ] Troubled, Disorganised	[ ] Rejuvenated, Generous	[ ] Progressed, Open	[ ] Shocked, Aware
[ ] Inexperienced, Uneducated	[ ] Contemplative, Cautious	[ ] Censored, Compromised	[ ] Meditative, Peaceful
[ ] Expectant, Apprehensive	[ ] Hindered, Provoked	[ ] Loyal, Dedicated	[ ] Developed, Awakened
[ ] Conflicted, Tense	[ ] Gracious, Idealistic	[ ] Opposed, Contradicted	[ ] Subordinate, Disadvantaged
[ ] United, Organised	[ ] Crest-fallen, Disabled	[ ] Obstructed, Threatened	[ ] Abundant, Accomplished
[ ] Sociable, Cooperative	[ ] Renewed, Optimistic	[ ] Liberated, Delivered	[ ] Mobile, Seeking
[ ] Restrained, Disappointed	[ ] Innocent, Truthful	[ ] Reduced, Impoverished	[ ] Gentle, Influential
[ ] Behavior-oriented, Self-aware	[ ] Strong, Vital	[ ] Advantaged, Beneficent	[ ] Joyous, Generous
[ ] Prosperous, Fruitful	[ ] Nurturant, Re-appraising	[ ] Resolute, Intentional	[ ] Fragmented, Ego-aware
[ ] Stagnant, Unassisted	[ ] Stressed, Challenged	[ ] Tempted, Seduced	[ ] Limited, Thrifty
[ ] Unselfish, Caring	[ ] Endangered, Unlucky	[ ] Community-oriented	[ ] Insightful, Unbiased
[ ] Supreme, Successful	[ ] Obligated, Dependent	[ ] Advanced, Fortunate	[ ] Conscientious, Conservative
[ ] Modest, Inhibited	[ ] Attractive, Liked	[ ] Oppressed, Exhausted	[ ] Balanced, Prospective
[ ] Enthusiastic, Harmonious	[ ] Steadfast, Matured	[ ] Wise, Hospitable	[ ] Hopeful, Reserved

APPENDIX B

INSTRUCTIONS TO JUDGES

You will be presented with a set of *I Ching* readings (64 in total). Readings for each hexagram are on a single page—there are 64 pages altogether (numbered at the bottom of each page in Roman numerals). A descriptor-pair and a rating scale (i.e., visual analogue scale; VAS) follows each reading (i.e., there is only one VAS to mark). You will read the reading, read the two-word descriptor in BOLD text underneath the reading, think about it, and then mark it in biro or pen for suitability on the VAS (see below). Do not spend too much time thinking about the content—vacillation and hesitation tend to impede one's judgment.

0% |_____| 100%
(not suitable) (suitable)

Your ratings will be <u>on</u> or <u>between</u>, 0% (not suitable) and 100% (suitable). By 'suitable' I mean the descriptor-pair accurately describes/summarises the content of the reading. 'Suitability' also carries the connotation that the descriptor-pair, on its own, sufficiently and adequately encapsulates the generic or overall message in the reading so there would be no confusion over intention between the descriptor-pair and its corresponding reading. In other words, you are **NOT** giving ratings on <u>accuracy</u> or <u>specificity</u> or <u>criticality</u>. . . . If you have any questions, do not hesitate to contact me for clarification. . . .

APPENDIX C

FREQUENCY TABLES

Table C1
Rankings of Descriptor-Pair Selections by Hexagram Number in Order of Preference (32 Least Preferred to 32 Most Preferred)

32 Least Preferred			32 Most Preferred		
Rank	Hexagram	Frequency	Rank	Hexagram	Frequency
64	23	56	32	15	179
63	54	57	31	58	192
62	14	61	29.5	26, 30	195
61	29	63	28	62	196
60	41	66	27	57	209
59	36	69	26	22	213
58	51	75	25	46	216
57	27	82	24	35	229
56	18	84	22	32, 34, 61	237
55	55	88	20	49	259
53.5	7, 33	95	19	5	264
52	11	98	16.5	13, 16, 25, 53	265
51	12	103	14	56	268
49.5	39, 40	105	13	24	277
48	21	107	12	31	281
47	38	119	11	6	285
46	45	125	10	3	303
44.5	4, 19	127	9	2	345
43	42	143	8	10	359
42	59	150	7	17	369
41	60	152	5.5	8, 37	371
40	50	153	4	20	379
38.5	9, 44	158	3	1	393
37	63	170	2	64	394
35.5	43, 52	175	1	28	465
34	48	176			
33	47	177			

APPENDIX C (CONT'D)

FREQUENCY TABLES

Table C2
First Hexagram Outcomes and Frequencies

Hexagram	Frequency
30	5
39, 42	7
24, 26, 40	8
1, 33, 38, 45, 59	9
10, 18, 28, 44, 58	10
6, 7, 13, 17, 27, 31, 50, 54, 55, 64	11
2, 16, 21, 23, 29, 35, 36, 48, 56, 60	12
3, 5, 15, 22, 62	13
14, 19, 37, 53, 57, 61	14
8, 9, 20, 25, 32, 41, 52	15
12, 46, 47, 49, 51	16
11, 43, 63	17
34	18
4	22

Table C3
Second Hexagram Outcomes and Frequencies

Hexagram	Frequency
46	3
41	5
13, 43, 60, 61, 64	6
1, 10, 19, 23, 37, 44	7
7, 9, 25, 27, 30, 36, 42, 45, 54, 55, 58	8
11, 14, 22, 24, 33, 35, 49, 52, 57, 62, 63	9
4, 6, 17, 32, 38, 50	10
3, 26, 34, 47, 48, 51, 56, 59	11
12, 16, 20, 40	12
28, 31, 39, 53	13
21, 29	14
2, 8	16
18	17
5	19
15	20