European Journal of Parapsychology

Volume 22.1, 2007

European Journal of Parapsychology

European Journal of Parapsychology

2007

Volume 22.1

INSTRUCTIONS TO AUTHORS

The European Journal of Parapsychology (EJP) is a peer-reviewed scientific journal for research - particularly theoretical and theory-driven empirical work - relating to the field of parapsychology (defined as the study of communication or interaction between organisms and their environment that do not appear to rely on the established sensorimotor channels). The Journal's aim is to stimulate and enhance activity in parapsychology, particularly in Europe, by publishing Articles, Research Notes, Reviews and Comments that offer insight into or criticisms of parapsychological research. The Journal also publishes peer-reviewed Student Research Briefs. Submissions are welcomed on any topic that falls under the heading of parapsychology as defined above. Authors are requested to briefly define any specific terms used (including 'standard' terms such as ESP, PK) in the text of the paper. All manuscripts submitted to the EJP must be under consideration solely by this journal and may not have been previously published in any form (with the exception of translations of non-English articles). Initial submissions should be sent via email to submissions@ejp.org.uk as an attached file (acceptable formats are ASCII, PDF, RTF, OpenOffice or MS Word).

ARTICLES are reports of experimental, clinical or theoretical research, with introductory comments and discussion of results generally being of direct relevance to the data or ideas being presented. Generally articles should be less than 10,000 words in length, though longer papers may be accepted after prior negotiation with the Editors. After evaluation by the editors, the initial article will be sent to two referees for peer-review.

RESEARCH NOTES are manuscripts of up to 3000 words in the following categories: conceptual replications; attempted replications which contradict the original study; reporting on new methodology or apparatus; new applications of existing methodology; challenges to key elements of an existing study, theory or methodology. Research Notes are not intended as a way of publishing inadequate or incomplete work. The Editors or Referees may also decide that a manuscript submitted as an article would be better published as a Research Note.

STUDENT RESEARCH BRIEFS are summaries of research (maximum 2000 words)conducted for assessment as part of a Bachelors or Masters degree (or equivalent) which would not otherwise be submitted as a full article but where the research is of sufficient interest to be disseminated to the wider parapsychological community. Student Research Briefs should provide a summary of methodology and results of the conducted research. They will be subject to peer-review with the understanding that the scope of the research will be less extensive than that of a full article.

REVIEWS provide critical and analytical review or assessment of previously published work. The EJP welcomes novel and cogent reviews (1500 words maximum), subject to Editorial approval.

COMMENTS are extended letters (1000 words maximum) that address specific issues raised in material published in the journal. All comments will be sent for peer-review, the process being similar to article review. Comments will be published in as timely a manner as possible, accompanied where appropriate with a reply from the original author(s).

Full formatting details are provided on the EJP website http://ejp.org.uk/

European Journal of Parapsychology

Implicit Learning in a Card Prediction Task

Stuart Wilson and Iain Hamlin
Queen Margaret University, Edinburgh,
United Kingdom

Abstract

Two experiments are described in which participants were asked to make predictions about which of four cards they would be presented with next. The order that the cards were presented in was governed by a non-salient underlying sequence, in one condition, and by a pseudo random-number generator in the other. In experiment one, playing cards were used, and no effects were found. In experiment two, Zener ESP cards were used and it was found that participants were more likely to make successful predictions in the sequence condition, but not the random condition. Furthermore, correct responses were associated with faster reaction times. We also found that extraversion was positively correlated with success in the sequence condition, as was the extent to which participants reported being "guided by psychic forces". These results are discussed in terms of "framing effects" and task demands.

Introduction

"Implicit learning" refers to the apparent ability of individuals to unconsciously pick up on subtle non-salient patterns in stimuli to which they are exposed. The term was coined by Arthur Reber in 1967, in reference to his finding that people could learn complex artificial grammars, despite being incapable of consciously articulating the grammatical rules. Since Reber's paper, a number of different paradigms have

Correspondence details: Stuart Wilson, Psychology, School of Social Sciences, Media and Communication, Queen Margaret University, Clerwood Terrace, Edinburgh, EH12 8TS. Email: swilson@qmu.ac.uk.

emerged, all purporting to demonstrate learning outside of conscious awareness. These include (but are not limited to): hidden covariation detection (e.g. Lewicki, 1986), contingent response tasks (e.g. . Buckalew, Finesmith and Sizemore, 1968), serial reaction time tasks (e.g. Nissen and Bullemer, 1987), control of dynamic systems tasks (e.g. Berry and Broadbent, 1987), invariant learning tasks (e.g. McGeorge and Burton, 1990). As a result of the explosion of different methodologies, the actual definition of what implicit learning is and how it is operationalised have become contentious issues in the field. There are at least a dozen different definitions (see Frensch, 1998) and the difficulties researchers have in agreeing what it actually means for learning to be implicit have led to numerous debates concerning whether or not the phenomenon actually exists (see Shanks and St. John, 1994).

Despite this, a standard definition of the phenomenon usually posits that implicit learning is learning in the absence of conscious awareness.[1] In one view of implicit learning, the knowledge gained is not accessible to consciousness, such that participants are unable to provide a verbal account of what they have learned.[2] Furthermore, the learning process itself does not involve conscious hypothesis testing, instead being an incidental consequence of the cognitive processing performed on the stimuli. So, implicit learning occurs when, after attending to a stimulus or series of stimuli, we unconsciously learn about the underlying properties of the stimulus/stimuli. These properties can take various forms, and this is reflected in the variety of implicit learning paradigms that have been developed.

The paradigm we wish to focus on is Implicit Sequence Learning (ISL). ISL tasks involve presenting participants with a series of stimuli, the order of which is governed by a particular pattern. During the course of the session, participants are asked to respond to the stimuli in some way, and this response usually indicates that they have become sensitive to the underlying rules governing the sequence. For example, in one kind of ISL task, participants are asked to respond to stimuli appearing at various locations on a computer screen. The order that the stimuli appear in follows a repeating sequence, and the learning is assessed by the time it takes participants to respond to each stimulus,

---

[1]Of course, what constitutes "conscious awareness" is one of the central debates in implicit learning literature

[2]The notion that verbal reports are sufficient measures of conscious awareness is itself fiercely contested by critics

with a faster reaction time suggesting that they had anticipated the location the stimulus would appear in. In other ISL tasks, participants are specifically asked to make predictions about the next stimulus that will be presented. It is this latter paradigm that we have adopted. For a full account of the issues surrounding ISL, see Cleeremans and Jimenez (1998).

Why does it matter? Implications for understanding anomalous experiences

In their day-to-day lives, humans are constantly engaged with a variety of complex systems. Many of these systems are rule-governed and, as a consequence, will exhibit subtle, recurring patterns or statistical regularities. It may be possible that some of the experiences reported as being "precognitive" or "intuitive" may be a result of individuals becoming sensitive to these patterns at an unconscious level, and then using this unconscious knowledge to make accurate predictions (as in implicit sequence learning). In the case of intuition, the knowledge may appear from nowhere and manifest as a "gut feeling" similar to the studies implementing "conceptual fluency" as a response modality. For example, imagine a man, John, who walks to work via the same route every day. He has been doing this for many years. On his way to work, he has to cross several busy roads, so must pay close attention to the traffic. It so happens that his route coincides with several bus-routes as they cross the city. One day, John gets caught in a particularly bad thunderstorm. Instinctively, he crosses the road and, after only a few seconds, a number 19 bus arrives, almost on cue. He boards the bus, which will take him straight to work, and he wonders what made him cross the road, how he knew the bus was due, and how he knew that the 19 was the appropriate bus to get.[3] It is possible that, during his years walking to work, he had unconsciously learned that the number 19 bus tended to stop in that location at roughly that point in his journey. Additionally, he unconsciously registered that the number 19 followed his walking route to work. To John, all that he was aware of was intuitively crossing the street, and somehow knowing that a bus would arrive to take him to work. With no conscious knowledge of how this information was obtained, it may seem to him mysterious, and depending on his worldview, may interpret it as being "psychic" "divine intervention" "intuition" etc. This is just one example, but there are other ways in which implicit learning may play a part in our day to day experience.

---

[3]Of course, this example is based on the premise that buses always run consistently on time.

The important point is that the unconscious nature of the phenomenon serves to mask the mechanism from the individual, thus leaving it open to interpretation.

Why does it matter? Implications for laboratory work

The importance of target-randomisation has been emphasised since the beginning of experimental parapsychology. Obviously, if the target or target-sequence is not randomised, then the possibility arises that participants may guess above chance based on unconscious knowledge of the non-random selection process. In other words, if the randomisation is inadequate, participants may learn (implicitly or otherwise) which target to expect.

Several critiques of experimental procedures in parapsychology have focused on the possibility that subtle sensory cues may be involved in producing above-chance results. For example, Hyman (1977) recognised that, in the original Ganzfeld procedure, using the same picture during judging as was handled by the "sender" may give participants a cue (again, either conscious or unconscious) as to its status as the "target". For example, the target picture may have fingerprints on it or may be at a slightly different temperature than the decoys. Although these features may not be picked up consciously (although this is not out of the question), they may be unconsciously noticed by the participant, giving them a tip-off that there is something different about that particular picture. Conceptually, this is in line with Lewicki's hidden covariation detection paradigm (Lewicki, 1986).

Implicit sequence learning has been linked with performance in forced-choice ESP studies in which participants are given trial-by-trial feedback. Gatlin (1977; 1979) has been strongly critical of techniques involving feedback, arguing that the degree of non-randomness in a target-sequence is positively correlated with above-chance scoring for that sequence. This suggests that it is the non-randomness that is driving the "hit-rate" rather than anything genuinely paranormal. This issue has been raised intermittently over the years, and the relevant issues are reviewed by Brugger and Taylor in a special edition of *The Journal of Consciousness Studies* (2003). Brugger and his colleagues contend that no finite sequence of target alternatives can be absolutely random and free of bias. As a result of this, any above-chance scoring on a psi test may be an artefact of the coherence between the way in which the target is selected and the responses made by participants. Brugger and Tay-

lor go further than Gatlin with their critique, suggesting that implicit learning may also play a role in forced-choice studies that do *not* give feedback. These authors suggest that all participants in a forced-choice study will display inherent biases in the responses that they make. If the target-sequence also contains a similar bias, then the result will either be above-chance performance (psi-hitting: when the two biases match) or below chance performance (psi-missing; when the two patterns are out of synch). Moreover, Brugger and Taylor also suggest that many of the individual difference factors that are thought to play a part in psi studies (e.g. the *sheep-goat* effect) are important in that they influence the biases displayed by the participants rather than influencing any purported psi ability.

Finally, implicit learning in a parapsychological context has also been demonstrated experimentally. Buckalew, Finesmith and Sisemore (1968) demonstrated that implicit learning could occur when participants taking part in a mock ESP test were presented with stimuli that conformed to a predetermined pattern. In this study, participants were asked to predict the suit of each card in a deck of standard playing cards, as held up by a fellow participant acting as the experimenter. Unknown to the participants, each deck of cards contained eight sequences of three cards each (heart, spade, diamond) sandwiched in with the rest of the deck randomly arranged around them. Buckalew et al. found that learning took place as evidenced by the increased number of "hits" as compared to chance. Additionally, no participants expressed any knowledge of the sequence when interviewed after the session.

While implicit learning has been extensively researched and debated in the domain of cognitive science (see, e.g. Shanks and St John, 1994), its practical applications, phenomenology and influence on human cognitive function (particularly problem solving and decision making) are areas that have been neglected to some extent. One main reason for this is the continued debate over some of the finer points of "implicit learning". Arguments concerning its nature (and indeed, its definition) have tended to dominate the literature. These debates are crucial if we are to understand such effects and what they can teach us about the wider issues concerning the mind, but they have also tended to obscure any "practical" applications for the observed (or at least, inferred) effect. Whilst we acknowledge that there are many unresolved issues concerning the mechanisms of such cognitive functioning (not least the question of how "implicit" such functions actually are, and how the knowledge is

represented; see Berry, 1997), we are basing our studies on the assumption that such learning effects do occur. What we are particularly interested in, therefore, are the ways in which such learning effects might contribute to the phenomena studied within parapsychology. Specifically, we are interested in the extent to which participants can utilise a repeating sequence when asked to make predictions based on their "intuition". We also wish to investigate the phenomenology behind this. Given that we are presenting the study within a particular frame, does this influence the way participants approach the task and their subsequent evaluation of their performance on the task?

Experiment one

Design

A repeated measures design was employed. Participants took part in two conditions, in which they viewed playing cards being presented on a computer screen, and were asked at various intervals to predict which card would come next. In the "implicit" condition, the sequence was governed by a non salient rule (see below). In the "random" condition, the order of the cards was determined by the computer's pseudo-RNG. This "random" condition may also be conceived as a traditional "psi" condition, as above chance performance in this condition may be interpreted as a psi effect by some researchers. The order in which participants took part in these two conditions was determined randomly.

Participants

61 participants[4] were recruited from students and staff of Queen Margaret University College and were paid £5 for their time.

Hypotheses

Primary hypotheses:

1. Predictions in the implicit learning condition will be above chance, when analysed in terms of "card colour".

2. "Psi believers" will perform better than "disbelievers".

---

[4]Unfortunately, due to a clerical error, the demographic details of the participants in both studies were accidentally destroyed.

3. Participants will score more "hits" in the implicit learning condition than the control/psi condition (based on the assumption that implicit learning is likely to be more prevalent and robust than psi).

Exploratory hypothesis:

1. There will be a correlation between performance in the implicit learning condition and in the psi condition (based on the idea that the same unconscious process mediates both abilities). This final hypothesis was exploratory.

Apparatus & materials

Two Flash MX programs were created to run the experiment on a Time 17″ Laptop. The programs displayed playing cards (Ace of Clubs, Ace of Spades, Ace of Hearts and Ace of Diamonds) to the participant, either randomly or governed by an underlying sequence.

For each participant, one of three repeating sequences of 16 was used to govern card order. The sequence itself was based on the colour of the card (red or black) rather than the suit of the cards. The suit, within each colour, of each card is determined at random by the program In order to create the sequences, 20 random sequences of 16 characters were taken from an internet-based random number generator. It was apparent that the vast majority of sequences contained too many long runs of characters (e.g. 0110000000001101) to be utilised, as participants would consciously pick up on the long runs after only a few repeats. The three sequences that were subsequently chosen were the best compromise we could make between long runs of single items and an alternating structure (e.g. 010101 etc). The three sequences were picked based on judgements by the researchers of how 'random' they would look to participants; 'random' sequences, as mentioned above, struck a happy medium between sequences with long runs and sequences of the alternating type. These three sequences were also balanced as to the number of "red" and "black" characters (i.e., 0s and 1s), so that a participant's systematic preference for any one colour (or card-type) would not lead to above or below chance performance.

The sequences employed in this experiment were as follows (b = black card, r = red card):

1. bbbrbbbrbrrbrrrbr

2. rbrrbbrbbbbrrbrr

3. rbrrrbrrbbbrbbrb

In the "random" condition, the sequence of the cards is determined randomly by the computer's random number generator. All other aspects are identical to the "implicit" condition.

A measure of psi-belief was also included in order to determine whether a person's belief in the possibility of psi effects impacted upon their performance. The measure we used was the Australian Sheep-Goat Questionnaire (18 item, forced choice version; Thalbourne and Delin, 1993).

Procedure

Participants were told that the experiment investigated the phenomenon of intuition. The researcher gave a short description of "intuitive" experiences, indicating that the information often appeared quickly and "out of the blue". Beyond this, intuition was not defined, and participants could interpret this as they wished. The experimenter asked participants not to dwell on their guesses, and also to avoid random guessing, allowing their 'intuition' to guide them. They were also asked to pay attention to all the stimuli and either to say each card quietly to themselves or to internally recite the cards as they were presented.

In the "implicit" condition, the program displayed a series of playing cards, each one appearing for 1 sec, and with a 0.5 sec gap between them. In order to aid participant's learning of the sequence, 32 cards (two cycles of the sequence) were displayed prior to the elicitation of the first guess.

After the second run-through of the sequence, the program started collecting "guesses". After a certain number (generated randomly by the computer with the constraint that it must be between 2 and 8) of cards were displayed, the program asked the participant *"which card will appear next?"* The participant was instructed to register their choice by clicking an on-screen symbol. Whether a guess was a "hit" or not was determined by the *colour* of the card they predict will be next. This is because the rule was based on the colour of the cards, and *not* the suit. So, if a participant predicted the Ace of Spades, and the actual card was Ace of Clubs, then this would be deemed a "hit", as the colour of the predicted card matched the colour of the next card in the sequence.

Forty guesses were made in each of the conditions (random and sequence). The procedure in the "random" condition was identical, with the exception that the order of the cards was determined by the computer's pseudo random number generator.

After the session was complete, the experimenter conducted an open-ended interview with the participant. During this interview, questions were asked concerning the participants' experience of the experiment. We were particularly interested in whether any participant had consciously detected any patterns during the experiment. We were also interested in any particular strategies that participants were using during the course of the session.

Results

Guess accuracy

The mean "correct colour" guesses were 19.87 (SD = 3.24) for the sequence condition, and 19.72 (SD = 2.86) for the random condition. The mean chance expectation for each condition was 20.

Neither guess accuracy in the random ($t_{(60)}$ = -0.32, p = .75, two tailed) or sequence condition ($t_{(60)}$ = 0.76, p = .45, two-tailed) differed significantly from chance. Guess accuracy in the sequence condition and the random condition did not significantly differ from each other ($t_{(59)}$ = 0.27, p = .79, two tailed).

The three sequences did not differ in terms of their "learnability" ($F_{(2,59)}$ = 1.22, p = .30), with subjects performing similarly in all three sequence conditions.

Effects of belief

"Sheep" and "Goats" were distinguished by means of a median-split. The scores on the Australian Sheep-Goat questionnaire ranged from a minimum of 2 to a maximum of 32 with a median of 18 (the lowest possible score on this questionnaire was 0, whilst the highest possible score was 36). This resulted in 27 participants being categorised as sheep and 34 being categorised as goats.

A one-way ANOVA was conducted in on order to determine whether belief influenced performance. No significant effect was found for either the random condition ($F_{(1,58)}$ = 0.58, p = .45) or the sequence condition ($F_{(1,59)}$ = 0.42, p = .52).

It became clear fairly quickly that most participants suspected that there might be a sequence underlying the order that the cards were displayed in. Many participants reported that they felt challenged to "figure out" what the experiment was about and tested various hypotheses, although none of these were close to the actual sequence. Indeed, many of these strategies were fairly crude (e.g. "every third card is a club") and did not reflect the complexity of the underlying sequence. Many participants expressed doubt that our experiment was actually on intuition. Indeed, due to the use of playing cards, some indicated that they thought the study was an investigation relating to gambling. One participant in particular expressed that his reservations about the study were due to the nature of the stimuli. In particular, he said that he doubted our "intuition" cover story as he specifically conceptualised intuition as being something "mysterious" and our study did not seem mysterious enough to be a legitimate test of intuition.

Discussion for experiment one

We did not find any significant implicit learning effect using this method. This may be because the sequences were too complex, or that participants were not attending to them sufficiently for implicit learning to take place.

It was clear in the post-session interviews that many participants suspected that there may be a pattern to the stimuli, and many of these participants actively sought to "solve" it. Obviously, doing this requires a great deal of conscious cognitive effort, as various hypotheses concerning what the sequence might be generated and tested by the participant during the course of the session. Such activity is unlikely to promote implicit learning, and this is potentially one reason why we found no effect in experiment one. Related to this, it is also notable that many participants doubted that we were actually looking at "intuition" as we had told them in our briefing session. This scepticism may also have had a part to play in the way participants approached the task. Indeed, some participants did indeed indicate that they spent time during the session attempting to work out what we were actually investigating. Again, this is unlikely to be favourable to any unconscious process that we were hoping to elicit.

Another alternative possibility is that the discrepancy between the colour-based rule and the participants' card-based guesses served to in-

hibit any real implicit learning effect. It is likely that participants were inclined to think in terms of card-suit. We had anticipated this, but hoped that, when deciding a card, the colour of the cards would have an unconscious influence (e.g. a participant might intuitively choose "ace of diamonds" because the colour-based sequence is unconsciously biasing them towards a red card, even if the actual target was the "ace of hearts"). In retrospect, looking at card-colour when participants are asked to guess card-suit may not have been successful, and the focus on the cards themselves may have over-ridden any subtle bias towards colour that the implicit learning might have provided.

Experiment two

A reaction time measure was employed in order to explore the possibility that participants exhibit implicit learning as indexed through faster reaction times on correct guesses than on incorrect guesses.

One participant's observation that our study did not seem "mysterious" enough to be a test of intuition got us thinking. Perhaps the playing cards that were used as stimuli in the first experiment serve to orient the participant in a particular direction. It may have been the case that the stimuli encouraged participants to approach the task in a particular frame of mind. As such, the nature of the stimuli may not have been conducive to implicit learning (or indeed psi). If participants are consciously attempting to "figure out" what is going on and attempting to develop strategies, then this may inhibit the passive conditions required for unconscious processing. In other words, consistent conscious deliberation may drown out any unconscious influence, be that implicit learning or psi.

As such, a further change implemented in experiment two was that Zener cards were used instead of playing cards in order to lend authenticity to the experiment. The use of widely known "ESP" cards would give credibility to the somewhat "mystical" nature of the experiment, whilst also being amenable to being used in a sequence experiment.[5] As a result of this, the new sequences were card-based rather than colour based. Colour-based sequences may actually have hampered participants' performance if they thought in terms of card-suit rather than

---

[5]Four cards were used instead of 5 in order to assist sequence creation. Had five cards been used, sequences would have contained more repeating segments and would have had to have been substantially longer to counterbalance this. As such, it was decided that sequences based on four cards would be used.

colour, because the rules consciously or unconsciously uncovered by participants during the trials would often be falsified by the fact that either one of two possible colour cards could be played at any point in the sequence (although whether this should matter if the effect is truly implicit, is debateable). Whereas, once again, there were 40 guesses in the random condition, this time there were 100 — as opposed to 40 — guesses in the sequence condition. It was hoped that this would give participants more time to learn the sequences and also add increased power to the analysis.

We also included the 'Big 5' personality dimensions as new variables in our experiment. Several commentators have noted the link between personality and performance in ESP tasks. Specifically, Honorton, Ferrari and Bem (1998), a meta-analysis, found a small significant positive correlation between extraversion and psi-performance on forced-choice studies (although the validity of this finding has been questioned due to heterogeneity and the fact that the correlation appeared to be dependent on whether participants knew their psi-score prior to completing the personality questionnaire). Woolhouse and Bayne (2000) have reported individual differences in performance on implicit learning tasks. The concept of their being individual differences in such tasks is a controversial one. The controversy can be illustrated by reference to Arthur Reber's thoughts on the issue. As previously noted, Reber is the pioneer of implicit learning research and has indicated that individual differences in implicit learning would be limited or non-existent (Reber, Walkenfeld and Hernstadt, 1991). According to Reber et al., this is because they consider implicit learning to be linked to evolutionarily older systems that are less susceptible to individual differences than systems relying on "later" conscious systems. However, by 2000 Reber appeared to have altered his opinion. In a paper with Rhianon Allen, he states "[i]t is our sense that individual differences in implicit functioning do exist. The question is whether such interindividual variation is distinguishable from what we commonly see in consciously modulated, top-down cognitive processing" (p. 242). Further research in this area is clearly needed. Woolhouse and Bayne (2002) gave people an implicit learning test of the "hidden covariation detection" type (see, e.g. Lewicki, 1986) in which participants were required to judge how suitable particular people were for a job based on personality profile scores. The suitability was determined by a non-salient rule concerning the scores. Woolhouse and Bayne report a significant relationship

between implicit learning scores and the sensing-intuition scale of the Myers-Briggs Type Indicator, with participants with a preference for intuition being more successful.

It would be interesting to see if the same dimensions of personality that affect implicit learning performance also affect psi performance, to see if these two phenomena are mediated by the same psychological variables. Although Woolhouse and Bayne used the MBTI, we decided to use a Big-5 questionnaire as our meaure of individual differences.

As in experiment one, we included the Australian Sheep-Goat questionnaire, and intended to see whether sheep and goats differed in terms of the strategies they reported adopting.

A further post-session questionnaire was administered to participants at the end of the experiment (see appendix). It investigated the psychological processes participants went through when making their guesses. Participants responded by circling a value (0%–100%) on an eleven point scale to indicate the frequency with which they went through each of the listed psychological processes when making their guesses. The items in the questionnaire represented the most frequently cited strategies from the debriefing interview in the previous study.

Participants

Forty-four participants were recruited via student/staff e-mails. All participants were students or staff members at Queen Margaret University College. Participants were paid £5 for their time.

Materials

The sequences utilised in this study were as follows:

1. waves, square, circle, circle, waves, cross, circle, square, cross, cross, circle, waves, square, square, cross, waves.

2. waves, square, square, circle, cross, waves, waves, circle, square, cross, cross, waves, cross, square, circle, circle.

3. waves, square, circle, cross, square, cross, waves, square, circle, circle, cross, cross, waves, waves, square, circle.

As before, we used the Australian Sheep-Goat Questionnaire to determine psi-belief and we measured personality using Goldberg's International Personality Item Pool.[6] Finally, we incorporated a post-session

---

[6]Scales derived from these markers show high convergent validity with the NEO (see Gow et al, 2005).

questionnaire based on comments made by participants in experiment one (see appendix). This covered strategies, perceptions of performance and attributions of performance.

Procedure

Participants were told, as in the previous experiment, that the experiment investigated the phenomenon of intuition. The researcher gave a short description of "intuitive" experiences, indicating that the information often appeared quickly and "out of the blue". The experimenter asked participants not to dwell on their guesses, and also to avoid random guessing, allowing their 'intuition' to guide them. Prior to taking part, participants completed the "Australian Sheep-Goat" questionnaire and the personality questionnaire.

As before, participants took part in both an implicit sequence condition (in which the order of the cards was governed by a pre-determined sequence) and a random (or psi) condition. Prior to beginning the session, the computer randomly selected which condition would be presented first.

In the "implicit" condition, the program displayed a series of Zener-ESP cards, each one appearing for 1 sec, and with a 0.5 sec gap between them. In order to aid participant's learning of the sequence, 32 cards (two cycles of the sequence) were displayed prior to the elicitation of the first guess. As indicated above, three sequences were employed, and which one of these was to serve as the sequence in a particular session was determined randomly by the computer.

After the second run-through of the sequence, the program started collecting "guesses". After a certain number (generated randomly by the computer with the constraint that it must be between 2 and 8) of cards were displayed, the program asked the participant *"which card will appear next?"* The participant was instructed to register their choice by clicking an on-screen symbol.

Forty guesses were made in the random condition, whilst 100 guesses were made in the sequence condition. The procedure in the "random" condition was identical, with the exception that the order of the cards was determined by the computer's pseudo random number generator. Reaction times were measured by the computer and were accurate to 1/10th of a second. After the session was complete, the participant was asked to complete another questionnaire (see appendix) relating to their experience of the task.

Results

The mean for the random condition was 9.86 (SD = 3.69), with a mean of 10 expected by chance. The mean for the sequence condition was 27.02 (SD = 6.81), with a mean of 25 expected by chance.

Performance in the sequence condition was marginally significant; $t_{(43)}$ = 1.97, p = .055 (2-tailed)[7]. Performance did not differ from chance in the random condition; $t_{(43)}$ = 0.25, p = .81).

Number of correct guesses in the random condition did not significantly correlate with number of correct guesses in the sequence condition (r = -.13, p = .93). This suggests that any psychological mechanisms helping participants perform well in one condition (i.e., sequence or random) do not help boost their performance in the other condition.

A comparison was also made between the random and sequence conditions in terms of the proportion of correct guesses made in each. No significant difference was found between the conditions; $t_{(42)}$ = 1.35, p = .19

Reaction time

We also looked at the reaction time data, to see if there was any difference between speed of reactions on implicit and control conditions. Mean reaction time was 2.14 seconds (SD = 0.57) for the random condition, and 2.25 seconds (SD = 0.67) for the sequence condition. No significant differences were found in the reaction times for responses in the random and sequence conditions ($t_{(42)}$ = 1.42, p = .16, two-tailed)

A binary logistic regression was performed to investigate whether or not time taken to make a guess was related to the accuracy of that guess in the sequence condition. A total of 4012 cases (individual guesses) were analysed. The model was significantly reliable (χ^2 = 3.96, df = 1, p = .05) in the sequence condition, suggesting that correct guesses were indeed made quicker than incorrect guesses. In the random condition, speed at which guesses were made was unrelated to guess accuracy (χ^2 = 0.20, df = 1, p = .65).

As in our previous experiment, there was no significant effect of sequence on number of correct guesses; $F_{(2,40)}$ = 0.72, p = .49. This suggests that all three sequences, despite differing in the extent to which they contain repeating sequence units, do not differ in terms of their

---

[7]It should be noted that we stuck to our previous convention of using two-tailed tests. Had this hypothesis been one-tailed, this effect would have been highly significant.

"learnability".

Individual differences and strategies

Correlational analyses were conducted to investigate the extent to which personality factors were associated with performance. Additionally, we correlated the extent to which various strategies and approaches were related to success in the respective tasks.

Table 1: Two-tailed Pearson's (r) correlations between successful performance, personality and reported strategy

Individual difference variable	Correct Random	Correct Sequence
Extraversion	.28	.43**
Agreeableness	-.003	.25
Conscientiousness	.23	-.15
Neuroticism	-.18	-.03
Openness	.18	.07
Extent to which participants stated they were:		
"guided by intuition"	.04	.21
"guided by psi"	.04	.34*
"first thing that comes into my head"	-.26	-.31
responding randomly	-.39*	-.05
looking for sequence	-.27	.13
influenced by infrequent cards	-.31	.15
influenced by particular card preference	-.26	-.16

*Correlation is significant at $p < .05$ level.
**Correlation is significant at $p < .01$ level.

Table 1 displays the correlations found between the various individual differences factors (including reported strategies) and accurate responses in the random and sequence conditions.

Extraversion did not positively correlate with number of correct guesses in the random condition, although how extraverted an individual was did seem to be correlated with performance in the sequence condition $r = .431, p = .008$. No other personality factor showed a significant correlation with performance in either condition.

Heuristics and strategies that are employed by participants could conceivably be related to how successful they are. The only reported strategy that correlated with performance in the random condition was a negative correlation between number of accurate responses and the extent to which participants responded randomly ($r = -.392, p = .02$).

The only reported strategy that correlated significantly with performance in the sequence condition was the extent to which participants reported being guided by "psychic forces" ($r = .344$, $p < .05$). It should be noted that, due to multiple analyses, the veracity of this marginally significant finding may be called into question. As these analyses relate to exploratory hypotheses, we have decided not to correct for multiple analyses, instead presenting the results and allowing the reader to evaluate them as to their significance.

Analyses were also conducted on the belief of individuals and the way this may have impacted upon performance and strategies. Data from two participants were discarded due to the belief questionnaire being incomplete. As before, the lowest possible score on this questionnaire was 0 (i.e., answered "false" to all belief questions) whilst the highest possible score was 36 (i.e., answered "true" to all belief questions). Data in the current study ranged from 2 to 30 with a median of 15.5. Participants falling below the median were classified as "goats", whilst those above the median were classified as "sheep". After splitting participants up in this way, there were 21 "sheep" and 21 "goats". Table 2 illustrates the performance of "sheep" and "goats" in each condition, as well as the mean scores for each group on the strategies they reported using.

Table 2: Means (and standard deviations) for performance indicators and reported strategies between "sheep" and "goats"

	Sheep		Goats	
Score in random ("psi") condition (chance = 10)	10.32	(3.9)	9.52	(3.89)
Score in sequence condition (chance = 25)	27.62	(7.39)	25.10	(4.88)
Extent to which participants stated they were:				
"guided by intuition"	38.68	(31.83)	23.61	(28.43)
"guided by psi"	23.42	(26.98)	5.83	(14.37)
"first thing that comes into my head"	40.53	(33.37)	48.16	(30.24)
responding randomly	21.58	(23.75)	27.63	(28.11)
looking for sequence	33.16	(37.37)	37.37	(28.45)
influenced by infrequent cards	27.5	(27.13)	28.42	(31.14)
influenced by particular card preference	33.42	(32.23)	28.16	(32.37)

An ANOVA was carried out with "belief" serving as the independent variable. Performance in both conditions, and reported strategies were dependent variables. No factors displayed a significant effect,

with the exception of "extent to which participants stated they were guided by 'psi'" ($F_{(1,36)}$ = 6.02, p = .02). Understandably, sheep reported that they were guided by psi significantly more than goats did.

Post-hoc analyses

In order to further explore the possibility that extraversion may have been associated with guessing strategy, post-hoc correlations were carried out on extraversion and the degree to which each participant reported using each strategy. From table 3, it can be seen that the only significant correlations were negative correlations between extraversion and the "extent to which participants stated that they were responding randomly" (r = -.44, p = .01) and the "extent to which participants stated that they were looking for a sequence" (r = -.33, p = .05).

Table 3: Two-tailed Pearson's (r) correlations between extraversion and reported strategies

Strategy	Correlation (r) with Extraversion
Extent to which participants stated they were:	
"guided by intuition"	.33
"guided by psi"	.22
"first thing that comes into my head"	-.26
responding randomly	-.44[**]
looking for sequence	-.33[*]
influenced by infrequent cards	-.05
influenced by particular card preference	.06

[*]Correlation is marginally significant at p = .05 level.
[**]Correlation is significant at p <.01 level.

Discussion for experiment two

Whereas in the previous study we failed to find any implicit learning effect, there were a number of interesting findings in the current experiment. Firstly, a marginally significant implicit learning effect was found, with participants guessing above chance when the cards were governed by an underlying sequence. In the control (psi) condition results were at chance levels. Furthermore, in the sequence condition it appeared that correct guesses were associated with a significantly reduced reaction-time, suggesting that the speed with which a response is made is a further indicator of implicit learning.

The success of the implicit learning condition can be interpreted in a number of ways. Firstly, it is possible that altering the stimuli

from playing cards (experiment one) to Zener cards (experiment two) changed the "frame" of the study. This may have led to participants approaching the task with a different frame of mind, perhaps one that was conducive to implicitly learning a non-salient sequence. It certainly seems as though the strategies employed by participants in experiment two were less concerned with attempts to consciously decipher what the study was about. Indeed, far fewer participants in the final study reported that they were actively seeking patterns in the stimuli. If it is the case that the act of searching for sequences actually inhibits the implicit sequence learning that might otherwise take place, then it is possible that the new methodology employed in experiment two served to promote implicit learning by letting participants attend to the stimuli without giving them reason to "work out" what was going on. This is further supported by the responses to the post-session questionnaire. An alternative explanation may be that the nature of the sequences in experiment one and experiment two was qualitatively different, and that it is this that lies behind the differing results in the two studies. Additionally, experiment two included a greater number of trials than the previous experiments. This may be important in order to elicit the subtle learning effects that we are interested in, as participants may require a greater number of trials for the learning to occur. Indeed, many other implicit learning experiments have a substantial number of sequence repetitions during one experimental session. Finally, it should also be remembered that this study was based on actual 1-to-1 "hits", rather than the colour based criteria implemented in the previous study. This may also have been an important factor. At the moment, it is not possible to determine which of these explanations is correct, and this is something that requires further work.

Regardless of the reason for the difference between experiments one and two, experiment two did independently produce evidence of implicit learning, both in terms of "correct guesses" and the reaction times associated with the correct guesses. When we look at the results of the post-session questionnaire, we begin to get an idea about what might have been driving participants' performance. The individual differences variable most strongly linked to performance in the sequence condition was extraversion; the higher the participant scored on extraversion, the more accurate they were in their guesses in the sequence condition ($r = .431$, $p < .01$). The extraversion finding should be considered in relation to Brugger and Taylor's suggestions. These authors sug-

gest that personality traits may influence "guessing behaviour" in such a way that participants with certain characteristics unwittingly manage to match their response patterns with underlying patterns in pseudo-random target sequences. So, extraverts score highly (say) on tests of ESP not because of any underlying psychic ability, but because they display certain features in their response patterning. From our study, it does seem that extraverts may be more sensitive to underlying patterns. It is also interesting to note the *post-hoc* negative correlation between extraversion and the extent to which participants reported that they were responding randomly. It seems that that extraverts are less likely to report responding randomly, which in turn may give them an advantage in learning the sequence. In addition to this, there was a marginally significant *post-hoc* finding suggesting a negative correlation between extraversion and the extent to which participants reported actively looking for a sequence. So, even though the extraverts did not seem to be responding randomly, they also did not report that they were consciously attempting to find the underlying rule. It may be the case that, for whatever reason, extraverts in our study were characterised by a tendency towards certain types of behaviour that enhanced their performance on the task. For example, avoiding responding randomly and avoiding actively looking for sequences suggests that these individuals may have been more comfortable with the demands made of them than other personalities. Thus, extraverts, when given a task of this kind, may be less self-conscious about getting "into the spirit" of the task. It is often overlooked how odd it must feel to participants when we ask them to "be psychic" or "use intuition". We tend to expect our participants to immediately know what to do in these circumstances. It is entirely possible that the way people approach the task itself and deal with the task-demands placed upon them is related to their personality profile. This is something that requires further attention, as there are clearly a number of contributing factors.

The only other factor that significantly correlated with performance in the sequence condition was the extent to which participants reported being guided by "psychic forces". It may appear somewhat strange that this correlation was significant, particularly in the sequence condition. However, the fact that participants who were successful in the sequence condition were those who reported being "guided by psychic forces" suggests that the best strategy for participants in this task was to "turn-off" their rational thinking strategies and allow "something else"

to take over. This "something else", however, was not merely intuition. The extent to which participants stated that they relied on their own intuition was not significantly correlated with successful performance. Instead, the most successful strategy seemed to be to approach the task with the view that "psi" was an influence in the responses. This may be due to the fact that, arguably, people conceptualise psi as being something external to themselves, whilst they see intuition as something that is more intrinsic to them. This may have been a factor in how active/passive participants were during the session, with those who let "psi do the work" the most passive (and thus the most amenable to implicitly learning the sequence, providing adequate attention was paid to the intervening trials). This is a matter for debate, however, and is worthy of further investigation. As a starting point, it would be interesting to see how locus-of-control (Rotter, 1966) interacts with both successful performance and strategy. This is something that has never really been addressed in the implicit learning literature, and could be a potentially fruitful area of research. It may be useful to also consider other personality constructs and how they might interact with the way people approach these tasks. As previously stated Woolhouse and Bayne (2000) used the Myers-Briggs Type Inventory and found that individuals who scored higher on the "sensing" dimension were more likely to perceive the underlying rule in an implicit learning task and then use it effectively. Woolhouse and Bayne also noted that individuals classified as "thinking" on the MBTI were more likely to use a deductive strategy that actually interfered with implicit learning performance, a finding which held true for individuals regardless of their score on the intuitive dimension. Woolhouse and Bayne do suggest that one possible reason why there was a difference between these personality types might be the strategy employed by them, although the authors only go so far as to suggest that this may be down to the way people approach the task in terms of the structure of the stimuli. It is clear from our study that the way people conceptualise what they are doing might also be a contributing factor. So, when the structure of the underlying pattern is essentially the same form, it is the nature of the stimuli and the way people conceptualise the task that interacts with their implicit learning performance. There are clearly many interacting factors to be teased apart.

Additionally, it is important for us to keep reminding ourselves of the potential applications of implicit learning for "real-world" scenar-

ios. Many implicit learning paradigms tend to be devoid of context and rather "cold". If implicit learning has a role to play in our everyday interactions, then we must also consider the variety of factors that may mediate the capacity to become sensitive to non-salient rule-governed stimuli.

The finding that correct guesses were associated with faster reaction times suggests that our implicit learning effect, although marginally significant, may be a valid one. This finding supports the notion that the effect is implicit in nature, as implicit processes tend to be thought of as being faster than explicit processes given their automaticity (e.g. see Hasher and Zacks, 1979, 1984). Thus, when participants "knew" the correct response (a result of implicit learning) they responded automatically, and thus faster than when they had no implicit knowledge. This increased speed of processing also supports the role of implicit learning in intuitive decisions, as, by definition, intuitive decisions are said to be faster than those involving explicit, rational processes.

In the random (or "psi") condition, we did not find any deviation from chance in terms of number of accurate guesses. Reaction times were not significantly related to accurate responses. The only factor that appeared to be related to correct guesses was the extent to which participants stated they were responding in a random fashion. This was a negative correlation suggesting that the less participants claimed to be responding randomly, the more successful they tended to be. Again, it is difficult to speculate why this might be the case, if it is a real effect. It may be that actively attempting to respond randomly interfered with any potential psi influence. Alternatively, it may be the case that, if participants noticed they were successfully guessing on a large number of trials, this may have subsequently influenced their response to the question concerning random responses (i.e., participants thought "I couldn't have been responding randomly if I was getting so many correct"). As there was no significant deviation from chance in the psi condition, it is likely that this effect reflects a chance finding.

For belief, we did not find any major differences between "sheep" and "goats". The only significant difference was in how much these groups reported being influenced by psi. It is no surprise to find that sheep reported this significantly more than goats, as those who don't believe in psi are not likely to report being guided by it.

General Discussion

We failed to find any implicit learning effect in experiment one, but did so in experiment two. It could be argued that that the different kind of stimuli used in the two studies encouraged participants to approach the tasks in different ways, with the Zener cards used in experiment two promoting a more detached approach, promoting implicit learning. This would make sense given the participants' accounts of their strategies in experiment one, with many participants indicating that they were consciously trying to "figure out" the experiment. This was not so widespread in experiment two. It is also possible that the differences between the sequences used in experiments one and two contributed to the differential pattern of performance between the two experiments. Exactly what might have been the underlying cause of the discrepancy between the two experiments is something that deserves further study.

If our results do indeed represent a framing effect in implicit learning, then questions may be asked about the role of task demands in implicit learning tasks. Many of the implicit learning tasks in the literature are less than engaging for the participant and lack any kind of context. For example, in the traditional artificial grammar learning paradigm (e.g. Reber, 1967) participants are presented with a series of letter strings that either conform to the grammar or do not (e.g. XMVTRXRM). In a serial reaction time experiment, participants press a button corresponding to the position of a light on a screen, the position of which is determined by a predetermined sequence. Some implicit learning paradigms do attempt to attach a context, but even in these cases, it is far from ecologically valid. For example, Berry and Broadbent (1988) had participants control the size of a workforce in an imaginary sugar-production factory, with the goal to maximise production. The production was governed by a rule (e.g. production = 2 × work-force – production on last trial). This attaches a context, but not one that is recognisable in the "real world". Tasks such as those described are useful when answering questions about the nature of implicit learning and the resulting issues concerning the way the information is processed and represented. Although the role of attention has been a popular topic in implicit learning research (see Hsiao and Reber, 1998), the nature of the demand characteristics and the motivation of the participants have never been considered. From the current research, it would appear that the expectations of the participant and the way in which they conceptualise the experiment

may be instrumental in guiding how they perform on that task. This may reflect how implicit learning might work in certain "real-world" settings. Additionally, it is important to consider the phenomenology of implicit learning and how it might be interpreted by individuals. The links with "intuitive" experiences are obvious, and it is not too much of a leap to suspect that implicit learning may have a part to play in the formation and maintenance of certain paranormal beliefs, if an individual is predisposed to interpret certain experiences in those terms.

Many of the points raised above have yet to be incorporated into mainstream implicit learning research. However, it may be time to develop a sideline in applied implicit learning. Its role in anomalous experience may prove to be one fruitful application, and it is hoped that more work will be conducted in this area.

Acknowledgements

References

Berry, D. C. (ed.) (1997) *How implicit is implicit learning?* Oxford University Press: Oxford.

Berry, D. C., and Broadbent, D. E. (1988) Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 36a,* 209-231.

Brugger, P. and Taylor, K. I. (2003) ESP: Extrasensory perception of effect of subjective probability? In J. Alcock, J. Burns and A. Freeman (Eds.) *Psi Wars: Getting to Grips with the Paranormal.* Imprint Academic Press: London. pp. 221-246.

Buckalew, L. W., Finesmith, S. H. and Sisemore, D. A. (1968) An experimental demonstration of learning without awareness. *Psychonomic Science, 11,* 277- 278.

Cleeremans, A. and Jimenez, L. (1998) Implicit sequence learning: The truth is in the details. In M.A. Stadler and P.A. Frensch (Eds.) *Handbook of implicit learning.* Sage Publications: London. pp. 323-364.

Frensch, P. A. (1998) One concept, multiple meanings: On how to define the concept of implicit learning. In M.A. Stadler and P.A. Frensch (Eds.) *Handbook of implicit learning.* Sage Publications: London. pp. 47-104.

Gatlin, L. L. (1977) Meaningful information creation: An alternative interpretation of the PSI phenomenon. *Journal of the American Society for Psychical Research, 71,* 1-18.

Gatlin, L.L (1979) A new measure of bias in finite sequences with application to ESP data. *Journal of the American Society for Psychical Research, 73,* 29-43.

Hasher, L., & Zacks, R. T. (1979) Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hasher, L., & Zacks, R. T. (1984) Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372-1388.

Honorton, C., Ferrari, D. C., Bem, D. J. (1998) Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology,*62, 255-276.

Hsiao, A. T. and Reber, A. (1998) The role of attention in implicit sequence learning: Exploring the limits of the cognitive unconscious. In M.A. Stadler and P.A. Frensch (Eds.) *Handbook of implicit learning.* Sage Publications: London. pp. 471-494.

Hyman, R. (1977). The case against parapsychology. *The Humanist, 37,* 37-49.

Lewicki, P. (1986) Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 135-146.

Lewicki, P., Czyzewska, M and Hoffman, H. (1987) Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 523-530.

McGeorge, P. and Burton, A. M. (1990). Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology, 42,* 597-610.

Nissen, M. J. and Bullemer, P. (1987) Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1-32.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855-863.

Reber, A. S., Walkenfeld, F. F and Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 88-96.

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In M. Stamenov (Series Ed.) & R.G. Kunzendorf & B. Wallace (Vol. Eds.) *Advances in consciousness research: Individual differences in conscious experience* pp. 228-247. Philadelphia: John Benjamins Publishing.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80,* whole issue.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367-447.

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

Woolhouse, L. S. and Bayne, R. (2000) Personality and the use of intuition: Individual differences in strategy and performance on an implicit learning task. *European Journal of Personality, 14,* 157-169.

Appendix

Debriefing Questions

Each of the questions was answered using the scale below following the instruction "Please circle the appropriate answer":

0...10...20...30...40...50...60...70...80...90...100% of the 140 trials.

General questions:

1. On what % of trials do you think you're guess was guided by something, psychic or otherwise?

2. On what % of trials do you think your guess was guided by something psychic?

Questions on "How did you make your guesses?":

1. You made your choice before the choice cards came up.

2. One of the cards in the flashing sequence struck you as prominent and influenced your choice of card.

3. One of the choice cards just looked right and so you picked it.

4. When the choice cards came up, you looked at all four, said their names in your head, then picked the one that sounded right.

5. You said the names of the flashing cards to yourself in your head as they flashed, and, when the choice card came up, you picked the card that *sounded* right.

6. You picked the first card that came into your head.

7. It came into your head as an image

8. It came into your head as a word

9. It came into your head as a feeling

10. A card came into your head, but you changed your mind and picked a different card.

11. A card came into your head, but you changed your mind and picked a different card instead AND THEN the card that you originally thought of turned out to be the correct card.

12. You picked randomly.

13. You felt compelled to pick the card, but just didn't know why.

14. You looked at the choice cards, ran the mouse over the cards and picked the one that

15. felt right.

16. You looked for a sequence in the flashing cards and picked a card based on this.

17. You picked a card you thought hadn't come up for a while.

18. Did you have a preference for a particular card (a preference that had nothing to do with psychic abilities) — e.g. were you particularly fond of the wavy lines card?

19. If you made your choices in some way not described here, please give details. Also, if you have any other comments, write them here.

On a Linear Assignment Permutation Test Applied to Parapsychological Data — Computational Enhancements and Additional Applications

Stephanie Stahl and Michael Brusco*

*Florida State University, Florida, USA

Abstract

Scott (1972) proposed a permutation test for the trace (sum of the diagonal) of a matrix that has received significant attention in the parapsychological literature. The matrix trace test has been used within the context of analyzing verbal and written material related to mediumship and remote viewing. We frame Scott's procedure within the broader context of linear assignment permutation tests. Using linear assignment, we demonstrate that the same matrix trace test has applicability to nominal and ordinal scale response agreement between two raters and to the Pearson product-moment correlation coefficient, both of which are used in parapsychological analyses. Along the way, we introduce a computational enhancement to facilitate the application of linear assignment exact permutation tests to matrices larger than those for which complete enumeration of all permutations is possible.

Introduction

Pratt and Birge (1948) developed a paradigm for testing mediumistic capability that has remained viable for more than one-half century (O'Keeffe & Wiseman, 2005; Scott, 1972). For this procedure, a medium (or mediums) produces a reading for each of n sitters. The

---

readings are broken into components and each sitter provides a rating of the accuracy of each component, with respect to themselves, for all readings. The components are totaled and the ratings are assembled in an $n \times n$ matrix where the rows correspond to the sitter for whom the reading was provided and the columns correspond to the sitter who produced the ratings. The diagonal elements of the matrix represent a one-to-one matching of each sitter with their ratings for their own reading. Therefore, the trace (or sum of the diagonal) of the matrix offers a measure of the sitters' ability to identify their own readings.

A formal statistical test associated with the $n \times n$ matrix of sitter ratings, which is attributed to Scott (1972), requires the evaluation of all $n!$ permutations of the columns of the matrix. For each permutation, the trace of the matrix is stored to create a reference distribution for the observed trace. A count of the number of permutations that yield a trace that is greater than or equal to the observed trace is obtained and divided by $n!$ to produce a one-tailed test of significance. For additional background on this test, we recommend the presentation offered by Utts (1993).

Scott's (1972) test has also received considerable usage in remote viewing experiments (Hansen, Utts, & Markwick, 1992; Schlitz & Gruber, 1980; Storm, 2003; Targ, 1994). In this context, remote viewers can provide a transcript for each of n target locations. A judge, who is blind with respect to which targets correspond to which transcripts, makes an assessment as to the accuracy of each transcript for each target, either on a holistic or a concept-by-concept basis. These assessments are tallied in an $n \times n$ matrix where the rows correspond to the target locations and columns correspond to the transcripts. The diagonal represents the concordance between each target and its corresponding transcript, and the significance test is the same as the one for mediumship.

Burdick and Kelly (1977, p. 109) suggested medium readings and remote target experiments as examples of free-response data. They described the advantage of the preferential rating technique over the forced-choice technique as a means to allow partial credit for resemblance of a response protocol to a target. For evaluating free-response material, Morris (1972) described the preferential matching exact test with respect to Stuart (1942). Solfvin, Kelly, and Burdick (1978) subsequently extended and generalized the work of Morris.

One of our objectives is to place Scott's (1972) procedure within the broader framework of linear assignment permutation tests, as devel-

oped by Hubert (1976, 1980, 1987, Chapter 2). We show that the same test of significance for the trace of a matrix proposed by Scott (1972) is directly applicable to other statistical problems relevant to parapsychology and related fields. Included among our applications are small sample tests of significance for the Pearson product-moment correlation coefficient (r) between two variables. When the two variables correspond to ranks, the Pearson correlation coefficient is equal to Spearman's rank correlation coefficient (ρ) and, therefore, the linear assignment test is related to this popular measure as well.

Another important application of linear assignment pertains to nominal and ordinal response scale agreement between two raters. Most notably, we develop the relationship between linear assignment and Cohen's (1960) kappa statistic, as well as a version of kappa applying appropriate weights to disagreements (Cohen, 1968; Fleiss, Levin, & Paik, 2003, Chapter 18). In the parapsychological literature, variations of kappa have been used in applications such as the measurement of agreement between two independent judges in Ganzfeld studies (Roe & Holt, 2004) and the concordance of two researchers coding published articles in academic literature (Watt & Nattegaal, 2004). Roe and Holt justified their preference for kappa over Spearman's rank correlation coefficient, with deference to Clark-Carter (1997, p. 533), because the latter measures the direction in which two sets of scores move rather than actual agreement. The weighted kappa is preferable for ordered categories, although its sensitivity to rater covariation renders it an absolute agreement measure relative to product-moment correlation (Schuster, 2004).

In addition to couching Scott's (1972) permutation test within the linear assignment paradigm and discussing related applications, we also focus on a new methodological development that enables exact tests to be obtained for somewhat larger matrices. We present results obtained using a new implicit enumeration procedure that performs exact tests without requiring explicit generation of all $n!$ permutations. We offer freely available programs in both Fortran and MatLab as well as an easy to use executable file, which parapsychological researchers can use to implement these procedures.[1]

In the following section, we present a formal description of the linear assignment paradigm, including a specification of the test statistic, a

---

[1]These programs can be downloaded free from: http://ejp.org.uk/index.php3?page=Download

discussion of generating the reference distribution via implicit enumeration, and a comparison of these approaches using a remote viewing matrix from the parapsychological literature. We subsequently devote a section to two other important applications of the linear assignment permutation test: (a) nominal and ordinal scale agreement between two raters, and (b) the Pearson correlation coefficient. We conclude with a brief summary.

Linear assignment and Scott's test

Formulation for linear assignment tests

A statistical test of the trace of an $n \times n$ matrix, **A**, can be characterized under Hubert's (1976, 1987, Chapter 2) paradigm of linear assignment. The testing process assumes that permutation (order) of the rows is fixed and that all $n!$ permutations of the columns are equally likely under the null hypothesis. Effectively, linear assignment refers to assigning each row object (label) to one and only one column object (label), creating a one-to-one correspondence (bijection) with the relevant information according to the assignment along the diagonal of the permuted matrix. Although the observed statistic is simple enough to calculate, $\sum_{i=1}^{n} a_{ii}$, the statistics must also be determined for all permutations of the columns to obtain the entire distribution in which the observed statistic lies. Using standard notation, Ψ is the set of all permutations of columns and, for each $\psi \in \Psi$, $\psi(i)$ is the object in position i of permutation ψ for $1 \leq i \leq n$. So, for the identity permutation (of columns), ψ_I, we have $\psi_I(i) = i$ for $i = 1,\ldots, n$, which can be written as $\psi_I = (1, 2, \ldots, n)$. A complete reference distribution can be mapped as follows:

$$\Gamma(\psi) = \sum_{i=1}^{n} a_{i\psi(i)} \qquad \text{for all } \psi \in \Psi. \tag{1}$$

The observed trace statistic as determined by the identity permutation, ψ_I, is:

$$\Gamma(\psi_I) = \sum_{i=1}^{n} a_{i\psi_I(i)}. \tag{2}$$

A one-tailed p-value for the significance of $\Gamma(\psi_I)$ can be performed by counting the number of permutations in the reference distribution (equation 1) with a trace index that is greater than or equal to the ob-

served statistic (equation 2) and dividing by the total number of possible permutations ($n!$).

Generation of the reference distribution

On current microcomputer platforms, the complete enumeration of all $n!$ permutations in the reference distribution is computationally feasible for $n \leq 14$. However, for as few as $n = 11$ or 12 objects, computation times can be measured in hundreds or thousands of seconds depending on the hardware and software platforms. We propose a computational procedure that greatly improves the efficiency of reference distribution generation for linear assignment tests. This procedure, which is based on concepts of branch-and-bound programming (Brusco & Stahl, 2005), *implicitly* generates the reference distribution without explicitly evaluating each and every permutation.

Our motivation for employing branch-and-bound methodology comes from various sources. Unfortunately, the number of permutations, $n!$, can be daunting when the object set is large. The required time to compute all of the statistics in the distribution can be quite lengthy or, possibly, infeasible. Mielke and Berry (2001, p. 2) present the three usual types of permutation tests as exact, resampling and moment approximation tests. When the likelihood of complete enumeration is low, the mean and variance of the distribution are often derived to facilitate a normal approximation for significance testing (Greville, 1944; Hubert, 1979; Hubert, 1987). In the early developmental stages of permutation testing, Wald and Wolfowitz (1944, p. 358) noted: "However, it is desirable to derive at least the limiting distributions of these statistics and make it practicable to carry out tests of significance when the sample is large." In response to the challenge posed by the size of the object set, randomisation tests are often employed, randomly choosing a sufficient number of permutations for evaluation to approximate the statistical distribution. However, a viable alternative to randomisation tests can be found in partial enumeration optimisation methodologies, yielding exact p-values rather than relying on approximation. Welch and Gutierrez (1988) applied branch-and-bound to exact significance testing within the context of matched-pairs designs. These authors noted: "Branch and bound is typically applied to optimisation problems, but the idea carries over to the p-value counting problem" (p. 451).

Our branch-and-bound procedure gradually builds permutations and evaluates the resulting partial permutations during the construc-

tion process. For example, consider a problem with $n = 15$ objects and a partial permutation of $p = 5$ of those objects, such as ($\psi(1) = 8$, $\psi(1) = 4$, $\psi(2) = 12$, $\psi(4) = 10$, $\psi(5) = 6$). If we can demonstrate that this partial permutation could only lead to a completed permutation with a trace index that equals or exceeds the observed statistic, then there is no reason to explicitly evaluate the $(n - p)! = 10! = 3{,}628{,}800$ complete permutations that stem from the partial permutation. Accordingly, we add 3,628,800 to the count of permutations with indices that equal or exceed the observed statistic. By similar argument, if we can show that the partial permutation cannot possibly lead to a completed permutation with a trace index that equals or exceeds the observed statistic, then we can also avoid explicit evaluation of the $(n - p)!$ permutations; however, in this case we add nothing to the permutation count. For our example, if either of these conditions ($> =$ or $<$) can be satisfied, then we redirect the building process to the next branch; specifically, we continue building permutations beginning with ($\psi(1) = 8$, $\psi(2) = 4$, $\psi(3) = 12$, $\psi(4) = 10$, $\psi(5) = 7$). In general, by using this pruning technique, we can often save ourselves the trouble of building and evaluating the $(n - p)!$ implicitly evaluated permutations. In effect, we begin with a program to build all permutations, but insert evaluation tests into the building process. The building process and relevant bound test are described in the appendix.

Numerical example

To demonstrate the linear assignment test and the computational savings afforded by the implicit generation of the reference distribution, we use a classic example from the remote viewing literature. The 10×10 matrix of scores displayed in table 1 presents transcontinental remote viewing data collected by Schlitz and Gruber (1980). We applied two MatLab programs that we developed for linear assignment tests to these data. The first program uses complete enumeration to generate the reference distribution, and finds that 17 of the $10! = 3{,}628{,}800$ column permutations of the data yield a trace greater than the observed trace. This results in a p-value of .000004684, which is the p-value reported by Schlitz and Gruber (1980). The complete enumeration algorithm required 97 seconds on a 2.2GHz Pentium IV PC.

The second MatLab program uses the implicit enumeration scheme. This program obtains the same p-value; however, the required CPU time was only 0.02 seconds. There is a reason that the implicit enumeration scheme is more than 4800 times faster. In particular, with

Table 1: Transcontinental remote viewing data from Schlitz and Gruber (1980, p. 311). The rows/columns (targets/responses) correspond to enumerated locations in the experiment.

	1	2	3	4	5	6	7	8	9	10
1	288	192	181	68	260	367	164	174	269	209
2	398	343	182	162	96	174	258	195	157	297
3	148	262	498	355	135	153	368	143	93	122
4	136	171	373	426	125	61	304	134	157	266
5	84	203	162	179	500	277	37	190	160	282
6	378	105	76	92	264	317	40	319	293	215
7	156	248	380	333	92	136	458	112	132	207
8	242	141	152	89	147	369	140	308	223	237
9	166	118	110	164	298	227	192	369	290	340
10	313	87	141	119	149	213	86	184	352	422

only 17 permutations with an index as good or better than the observed statistic, the implicit enumeration program quickly determines that there is no possible way that many partial permutations can be completed to meet or exceed the observed statistic. The rapid execution of our illustration demonstrates a fundamental characteristic of the efficiency of the partial enumeration strategy. In particular, *early* pruning (for partial solutions consisting of only a few objects) is likely when the observed statistic is truly significant within the distribution generated by equation (1) and when we have tight bounding procedures (reasonable upper and lower bounds for the $>=$ and $<$ tests). Specifically, the algorithm performs more quickly when the p-value is very small. In other words, this approach is most desirable for analysts who have high confidence in their hypotheses.

Other applications of linear assignment

Classification agreement between two raters

Cohen's (1960) kappa (κ) statistic is a well-recognized measure of the agreement between two raters who are asked to classify N objects (o_1, o_2, ..., o_N) into T nominal (i.e., unordered) groups (G_1, G_2, ..., G_T). A common representation of data collected from the raters is a $T \times T$ contingency table with elements n_{ij} representing the number of objects that are assigned by rater 1 to group G_i and by rater 2 to group G_j. The contingency table has the following structure, where $n_i.$ is the sum of row i ($1 \leq i \leq T$) and $n_{.j}$ is the sum of column, j, which is shown in

table 2.

Table 2: Contingency table for two raters classifying N objects into T groups

		Rater 2				Row Sum
		G_1	G_2	...	G_T	
	G_1	n_{11}	n_{12}	...	n_{1T}	$n_1.$
	G_2	n_{21}	n_{22}	...	n_{2T}	$n_2.$
Rater 1	...	...	...	...	...	...
	...	...	...	...	...	...
	G_T	n_{T1}	n_{T2}	...	n_{TT}	$n_T.$
	Column Sum	$n._1$	$n._2$	...	$n._T$	n

Usually denoted to compare observed and expected outcomes as $\kappa = \frac{\pi_O - \pi_E}{1 - \pi_E}$ in terms of probabilities, Cohen's measure of nominal scale agreement index is computed from the data in the contingency table using the formula:

$$\kappa = \frac{\left[\left(\frac{1}{n}\sum_{i=1}^{T} n_{ii}\right) - \left(\frac{1}{n^2}\sum_{i=1}^{T} n_{i\bullet}n_{\bullet i}\right)\right]}{\left[1 - \left(\frac{1}{n^2}\sum_{i=1}^{T} n_{i\bullet}n_{\bullet i}\right)\right]} \quad (3)$$

The term to the left of the minus sign in the numerator of equation (3) is π_O, the observed proportion of agreement between the two raters. The term to the right of the minus sign in the numerator is π_E, the expected proportion of agreement. Because the denominator is one minus the expected agreement proportion, Cohen's κ represents the proportion of agreement between the two raters after allowing for chance.

Cohen (1968) provided a generalisation of κ to allow for "partial credit" in the agreement between two raters. This index, which is known as weighted kappa (κ_W), is particularly appropriate when the groups exhibit an ordinal relationship (i.e., the groups are ordered). For example, suppose that rater 1 placed object $o1$ in group G_1 and rater 2 classified object $o1$ in G_2, then some partial credit might be afforded for object $o1$ because the classifications of the two raters differs by only one group position. Less credit would be granted if rater 2 placed object $o1$ in G_3 because this placement is an additional group position from rater 1's classification of the same object. Schuster (2004) explains how weighted kappa is a good measure when one rater consistently gives higher (more optimistic) ratings than the other.

The most common weighting schemes assume unit distance between the ordered groups with either linear or quadratic weights. The linear weights are usually determined by absolute values of disagreements and the quadratic weights by squares of disagreements (Agresti, 2002, p. 430; Fleiss et al., 2003, Chapter 18). For example, Cicchetti and Allison (1971) proposed the following weights for each cell of the contingency table based on a linear scheme:

$$w_{ij} = 1 - \left(\frac{\mid i - j \mid}{(T-1)} \right), \qquad \text{for } 1 \leq i, j \leq T. \tag{4}$$

Similarly, Fleiss and Cohen (1973) proposed the following quadratic weighting scheme:

$$w_{ij} = 1 - \left(\frac{\mid i - j \mid^2}{(T-1)^2} \right), \qquad \text{for } 1 \leq i, j \leq T. \tag{5}$$

The quadratic weighting scheme of equation (5) is especially important because it has a strong relationship with other measures of association. Schuster (2004), for example, demonstrates that weighted kappa based on equation (5) is equal to an intraclass correlation coefficient when the means of the two raters are equal, and equal to the Pearson product moment correlation coefficient, r, when both the means and variances of the raters are equal.

For any selected weighting scheme, Cohen's (1968) weighted kappa index is computed as follows:

$$\kappa_W = \frac{\left[\left(\frac{1}{n} \sum\limits_{i=1}^{T} \sum\limits_{j=1}^{T} w_{ij} n_{ij} \right) - \left(\frac{1}{n^2} \sum\limits_{i=1}^{T} \sum\limits_{j=1}^{T} w_{ij} n_{i\bullet} n_{\bullet j} \right) \right]}{\left[1 - \left(\frac{1}{n^2} \sum\limits_{i=1}^{T} \sum\limits_{j=1}^{T} w_{ij} n_{i\bullet} n_{\bullet j} \right) \right]}. \tag{6}$$

We note that if $w_{ij} = 1$ for all $1 \leq i = j \leq T$ and $w_{ij} = 0$ for all $1 \leq i \neq j \leq T$, then equation (6) reduces to equation (3).

Hubert (1976, 1980) provides an elegant representation of weighted kappa that serves as the basis of the equivalence between rater agreement and Scott's (1972) matrix trace test. In particular, Hubert defines a series of matrices, $\mathbf{U}_{ij}$ (for $1 \leq i, j \leq T$), of dimension $n_{i\cdot} \times n_{\cdot j}$, where each element of $\mathbf{U}_{ij}$ is equal to w_{ij}. An $n \times n$ matrix $\mathbf{C}$, is displayed below:

$$\mathbf{C} = \begin{vmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} & \dots & \mathbf{U}_{1T} \\ \mathbf{U}_{21} & \mathbf{U}_{22} & \dots & \mathbf{U}_{2T} \\ \dots & \dots & \dots & \dots \\ \mathbf{U}_{T1} & \mathbf{U}_{T2} & \dots & \mathbf{U}_{TT} \end{vmatrix}$$

For convenience and without loss of generality, Hubert (1980) assumes that rows of $\mathbf{C}$, which correspond to rater 1's classification of objects, have been labeled in order, such that objects 1 to $n_1.$ are in the first $n_1.$ rows, objects $(n_1. + 1)$ to $(n_1. + n_2.)$ are in the next $n_2.$ rows, etc. The object labels for the columns, however, will typically not be the same as the object labels for the rows. For example, the labels for the first $n._1$ columns must correspond to those objects that were classified by rater 2 in G_1. Notice that the *values* in $\mathbf{C}$ are the weights, but the arrangement of the weights depends on the actual data. Hubert then defines an $n \times n$ permutation matrix, $\mathbf{D}$, with elements $d_{ij} = 1$ if the object label for row i equals the object label for column j and 0 otherwise.

To illustrate the matrix trace test (i.e., linear assignment) for rater agreement, we use data published by Rae (1996, p. 841) that assumes two clinicians classify $N = 10$ patients into one of $T = 3$ diagnostic categories: G_1 = personality disorder, G_2 = neurosis, or G_3 = psychosis. The results of the classification process are shown in table 3, and the corresponding contingency table is provided in table 4.

Table 3: Data for rater agreement from Rae (1996, p. 841) after reordering the patients for convenience of presentation

Patient	1	2	3	4	5	6	7	8	9	10
Clinician 1	G_1	G_1	G_2	G_2	G_2	G_2	G_3	G_3	G_3	G_3
Clinician 2	G_1	G_1	G_1	G_2	G_2	G_3	G_1	G_1	G_2	G_3

Table 4: Contingency table for rater agreement data in Table 3

		Clinician 2			
		G_1	G_2	G_3	*Row Sum*
	G_1	2	0	0	2
Clinician 1	G_2	1	2	1	4
	G_3	2	1	1	4
	Column Sum	5	3	2	10

We consider measurement of rater agreement via kappa, as well as weighted kappa for the linear and quadratic weighting schemes. The computation of kappa using equation (3) is as follows:

$$\kappa = \frac{\left[\left(\frac{1}{10}(2+2+1)\right) - \left(\frac{1}{100}\left((2)(5)+(4)(3)+(4)(2)\right)\right)\right]}{\left[1 - \left(\frac{1}{100}\left((2)(5)+(4)(3)+(4)(2)\right)\right)\right]} = 0.2857. \quad (7)$$

For a linear weighting scheme, the appropriate weight matrix, $\mathbf{W}_L$, is shown in table 5. Using these weights in equation (6), the computation of weighted kappa yields $\kappa_W = 0.2553$. For the quadratic weighting scheme, the appropriate weight matrix, $\mathbf{W}_Q$, is displayed in table 6 and the computation of weighted kappa yields $\kappa_W = 0.2254$.

Table 5: Weighted matrix, $\mathbf{W}_L$, for a linear weighting scheme, equation (4), when $T = 3$

	G_1	G_2	G_3
G_1	1.00	0.50	0.00
G_2	0.50	1.00	0.50
G_3	0.00	0.50	1.00

Table 6: Weighted matrix, $\mathbf{W}_Q$, for a quadratic weighting scheme, equation (5), when $T = 3$

	G_1	G_2	G_3
G_1	1.00	0.75	0.00
G_2	0.75	1.00	0.75
G_3	0.00	0.75	1.00

For the sake of parsimony, we provide a detailed description of the linear assignment test for rater agreement only within the context of the quadratic weighting scheme. Matrices $\mathbf{C}$ and $\mathbf{D}$ are displayed in the top and bottom panels of table 7, respectively. Multiplication of the matrices in the top and bottom panels of table 7 produces the matrix $\mathbf{CD}^T$, which is displayed in table 8.

The trace of matrix $\mathbf{CD}^T$ is 7.25. The observed weighted agreement can also be directly computed from the contingency table and $\mathbf{W}_Q$ as:

$$\Gamma(\psi_I) = \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ij} n_{ij} = 2(1.00) + 1(0.75)$$
$$+ 2(1.00) + 1(0.75) + 1(0.75) + 1(1.00) = 7.25$$

Using the MatLab implementation of the linear assignment test based on exhaustive enumeration of all 10! permutations of the columns

Table 7: Matrices **C** (top panel) and **D** (bottom panel) for rater agreement assuming quadratic weighting scheme

	1	2	3	7	8	4	5	9	6	10
1	**1.00**	1.00	1.00	1.00	1.00	0.75	0.75	0.75	0.00	0.00
2	1.00	**1.00**	1.00	1.00	1.00	0.75	0.75	0.75	0.00	0.00
3	0.75	0.75	**0.75**	0.75	0.75	1.00	1.00	1.00	0.75	0.75
4	0.75	0.75	0.75	0.75	0.75	**1.00**	1.00	1.00	0.75	0.75
5	0.75	0.75	0.75	0.75	0.75	1.00	**1.00**	1.00	0.75	0.75
6	0.75	0.75	0.75	0.75	0.75	1.00	1.00	1.00	**0.75**	0.75
7	0.00	0.00	0.00	**0.00**	0.00	0.75	0.75	0.75	1.00	1.00
8	0.00	0.00	0.00	0.00	**0.00**	0.75	0.75	0.75	1.00	1.00
9	0.00	0.00	0.00	0.00	0.00	0.75	0.75	**0.75**	1.00	1.00
10	0.00	0.00	0.00	0.00	0.00	0.75	0.75	0.75	1.00	**1.00**
1	**1**	0	0	0	0	0	0	0	0	0
2	0	**1**	0	0	0	0	0	0	0	0
3	0	0	**1**	0	0	0	0	0	0	0
4	0	0	0	0	0	**1**	0	0	0	0
5	0	0	0	0	0	0	**1**	0	0	0
6	0	0	0	0	0	0	0	0	**1**	0
7	0	0	0	**1**	0	0	0	0	0	0
8	0	0	0	0	**1**	0	0	0	0	0
9	0	0	0	0	0	0	0	**1**	0	0
10	0	0	0	0	0	0	0	0	0	**1**

Table 8: Matrix $\mathbf{CD}^T$ for rater agreement assuming quadratic weighting scheme

	1	2	3	4	5	6	7	8	9	10
1	1.00	1.00	1.00	0.75	0.75	0.00	1.00	1.00	0.75	0.00
2	1.00	1.00	1.00	0.75	0.75	0.00	1.00	1.00	0.75	0.00
3	0.75	0.75	0.75	1.00	1.00	0.75	0.75	0.75	1.00	0.75
4	0.75	0.75	0.75	1.00	1.00	0.75	0.75	0.75	1.00	0.75
5	0.75	0.75	0.75	1.00	1.00	0.75	0.75	0.75	1.00	0.75
6	0.75	0.75	0.75	1.00	1.00	0.75	0.75	0.75	1.00	0.75
7	0.00	0.00	0.00	0.75	0.75	1.00	0.00	0.00	0.75	1.00
8	0.00	0.00	0.00	0.75	0.75	1.00	0.00	0.00	0.75	1.00
9	0.00	0.00	0.00	0.75	0.75	1.00	0.00	0.00	0.75	1.00
10	0.00	0.00	0.00	0.75	0.75	1.00	0.00	0.00	0.75	1.00

of $\mathbf{CD}^T$, we obtained an exact p-value of .29524 for rater agreement under the quadratic weighting scheme. The required computation time for this process was 90.14 seconds of CPU time. Using the implicit enumeration method, the same p-value was obtained in only 5.56 seconds.

We also used the MatLab programs to test rater agreement under the assumption of a linear weighting scheme, as well as for the unweighted kappa index. For the unweighted index, where only exact matches are credited, the exact p-value of .15238 was obtained in 90.19 seconds using complete enumeration, but only 3.17 seconds using implicit enumeration. Under the linear weighting scheme, the exact p-value of .24603 was obtained in 90.23 and 5.20 seconds using complete enumeration and implicit enumeration, respectively. We note that the same exact p-values for the unweighted and weighted kappa indices can also be obtained much more efficiently by applying a permutation test to the contingency table, while maintaining fixed marginal values for the rows and columns (see Berry, Johnston, & Mielke, 2005). Thus, we are not tacitly advocating that Scott's (1972) matrix trace approach is the best way to proceed for testing weighted kappa. Instead, our goal was merely to show the matrix trace test could be used for this purpose. For implementation purposes, the methodology can be used for multiple statistical tools within a single program.

Pearson product-moment correlation coefficient

The linear assignment test can also be used to conduct a test of significance for the Pearson correlation coefficient between two variables (Hubert, 1976). Moreover, as the Spearman correlation coefficient is equal to the Pearson correlation coefficient when the two variables are ranks, the test is also relevant to the Spearman correlation measure, which frequently appears in the parapsychological literature. To illustrate a test of the Pearson correlation coefficient, let x_i and y_i denote measurements on two variables for each of n observations, $1 \leq i \leq n$, and define $\bar{x}$ and $\bar{y}$ as the means for these two variables, respectively. We then define the following $n \times n$ matrix:

$$\mathbf{A} = \{a_{ij} = (x_i - \bar{x})(y_j - \bar{y})\} =$$

$$\begin{bmatrix} (x_1 - \bar{x})(y_1 - \bar{y}) & (x_1 - \bar{x})(y_2 - \bar{y}) & \ldots & (x_1 - \bar{x})(y_n - \bar{y}) \\ (x_2 - \bar{x})(y_1 - \bar{y}) & (x_2 - \bar{x})(y_2 - \bar{y}) & \ldots & (x_2 - \bar{x})(y_n - \bar{y}) \\ \ldots & \ldots & \ldots & \ldots \\ (x_n - \bar{x})(y_1 - \bar{y}) & (x_n - \bar{x})(y_2 - \bar{y}) & \ldots & (x_n - \bar{x})(y_n - \bar{y}) \end{bmatrix}$$

Consider the formula for the Pearson correlation coefficient (Groebner, Shannon, Fry, & Smith, 2001, p. 427):

$$r = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum\limits_{i=1}^{n} (x_i - \bar{x})^2\right)\left(\sum\limits_{i=1}^{n} (y_i - \bar{y})^2\right)}} \tag{8}$$

The numerator of the formula for r is equal to the trace of $\mathbf{A}$, and the denominator is invariant to all $n!$ relabelings of the columns of $\mathbf{A}$. Regarding the use of the linear assignment test for correlation, we use the diagnostic assignments of clinicians 1 and 2 from table 3 as the x and y variables, respectively. The raw data for the correlation analysis are displayed in table 9, and matrix A for the linear assignment test is shown in table 10.

Table 9: Variables for the correlation example, where x_i (y_i) corresponds to the group assignments provided by clinician 1 (clinician 2) for the 10 patients, as shown in table 4

Patient	1	2	3	4	5	6	7	8	9	10
x_i	1	1	2	2	2	2	3	3	3	3
y_i	1	1	1	2	2	3	1	1	2	3

Table 10: Matrix $\mathbf{A}$ for the correlation example

	1	2	3	4	5	6	7	8	9	10
1	0.84	0.84	0.84	-0.36	-0.36	-1.56	0.84	0.84	-0.36	-1.56
2	0.84	0.84	0.84	-0.36	-0.36	-1.56	0.84	0.84	-0.36	-1.56
3	0.14	0.14	0.14	-0.06	-0.06	-0.26	0.14	0.14	-0.06	-0.26
4	0.14	0.14	0.14	-0.06	-0.06	-0.26	0.14	0.14	-0.06	-0.26
5	0.14	0.14	0.14	-0.06	-0.06	-0.26	0.14	0.14	-0.06	-0.26
6	0.14	0.14	0.14	-0.06	-0.06	-0.26	0.14	0.14	-0.06	-0.26
7	-0.56	-0.56	-0.56	0.24	0.24	1.04	-0.56	-0.56	0.24	1.04
8	-0.56	-0.56	-0.56	0.24	0.24	1.04	-0.56	-0.56	0.24	1.04
9	-0.56	-0.56	-0.56	0.24	0.24	1.04	-0.56	-0.56	0.24	1.04
10	-0.56	-0.56	-0.56	0.24	0.24	1.04	-0.56	-0.56	0.24	1.04

The correlation between the two variables is $r = .2738$, with a t-statistic of 0.8050 and a one-tailed p-value of .22204 based on a normal approximation. We ran a MatLab implementation of the exact permutation test for correlation described by Edgington (1995, Chapter 9), which is based on complete enumeration of the reference distribution. This algorithm required 101.65 seconds and yielded a one-tailed p-value of

.29524. The implicit enumeration algorithm applied to **A** yielded the same p-value in 4.06 seconds of CPU time. In short, the first p-value is approximate, and the Edgington p-value is exact but takes much longer to execute than our method that calculates the same exact p-value.

Summary and discussion

Permutation tests play a prominent role in a number of areas of parapsychological research (Radin, Machado, & Zangari, 1998; Schmidt, Schneider, Binder, Burkle, & Walach, 2001; Stahl, 2004; Targ, 1994; Utts, 1989, 1993). In this paper, we have focused on permutation tests related to the linear assignment model, which encompasses Scott's (1972) well-known procedure that has been used in mediumship and remote viewing research. We also discuss applications of the linear assignment paradigm to other important classes of problems, emphasizing rater agreement and correlation.

We offer freely available software programs in two languages that can be used to test the significance of the trace of an $n \times n$ matrix. Most importantly, these are programs that use implicit rather than explicit generation of all permutations, and thus produce exact tests in a fraction of the time required for complete generation of all permutations.

The extension of the linear assignment paradigm to quadratic assignment will offer more tests to the parapsychological researcher's quantitative toolbox. Stahl (2004) recently discussed several of these tests, which include Mantel's (1967) test for the agreement of two matrices, symmetry tests, and tests of within-row and/or column patterning of two matrices. Whereas linear assignment tests require permutations of the rows *or* columns of the $n \times n$ matrix, the quadratic assignment tests described by Hubert (1987, Chapters 4 and 5) and Stahl require congruent permutation of the rows *and* columns. Research is currently underway to develop the implicit generation procedure for the more challenging area of quadratic assignment tests.

References

Agresti, A. (2002). *Categorical data analysis* (2nd Ed.). New York: Wiley.

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2005). Exact and resampling probability values for weighted kappa. *Psychological Reports, 96*, 243-252.

Brusco, M. J., & Stahl, S. (2005). *Branch-and-bound applications in combinatorial data analysis*. New York: Springer.

Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In B. B. Wolman (Ed.) *Handbook of parapsychology*. New York: Van Nostrand Reinhold, pp. 81-130.

Cicchetti, D. V., & Allison, T. (1971). A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology, 11*, 101-109.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213-220.

Edgington, E. S. (1995). *Randomization tests* (3rd ed.). New York: Marcel Dekker.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions.* (3rd ed.). New York: Wiley.

Greville, T. N. E. (1944). On multiple matching with one variable deck. *The Annals of Mathematical Statistics, 15*, 432-434.

Groebner, D. F., Shannon, P. W., Fry, P. C., & Smith, K. D. (2001). *Business statistics: A decision-making approach.* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hansen, G. P., Utts, J. And Markwick, B. (1992). Critique of PEAR remote-viewing experiments. *Journal of Parapsychology, 56*, 97-114.

Hubert, L. J. (1976). A relationship between the assignment problem and some simple statistical techniques. *Quality and Quantity, 10*, 341-348.

Hubert, L. J. (1979). Matching models in the analysis of cross-classifications. *Psychometrika, 44*, 21-41.

Hubert, L. J. (1980). Alternative inference models based on matching for a weighted index of nominal scale response agreement. *Quality and Quantity, 14*, 711-725.

Hubert, L. J. (1987) *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27*, 209-220.

Mielke, P. W., & Berry, K. J. (2001). *Permutation Methods: A distance function approach*, New York: Springer.

Morris, R. L. (1972). An exact method for evaluating preferentially matched free-response data. *Journal of the American Society for Psychical Research, 66*, 401-407.

O'Keeffe, C., & Wiseman, R. (2005). Testing alleged mediumship: methods and results. *British Journal of Psychology, 96*, 165-179.

Pratt, J. G., & Birge, W. R. (1948). Appraising verbal test material in parapsychology. *Journal of Parapsychology, 12*, 236-256.

Radin, D. I., Machado, F. R., & Zangari, W. (1998). Effects of distant healing intention through time and space: Two exploratory studies. *Proceedings of Presented Papers: The Parapsychological Association 41st Annual Convention*, 143-161.

Rae, G. (1996). Exact probabilities for the general matching paradigm. *Educational and Psychological Measurement, 56*, 839-842.

Roe, C. A., & Holt, N. (2004). A further consideration of the sender as a PK agent in Ganzfeld ESP studies. *Proceedings of Presented Papers: The Parapsychological Associ-*

ation 47th Annual Convention, 197-206.

Schlitz, M., & Gruber, E. (1980). Transcontinental remote viewing. *Journal of Parapsychology, 44*, 305-317.

Schmidt, S., Schneider, R., Binder, M., Bürkle, D., & Walach, H. (2001). Investigating methodological issues in EDA-DMILS: Results from a Pilot Study. *Journal of Parapsychology, 65*, 59-82.

Schuster, C. (2004). A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. *Educational and Psychological Measurement, 64*, 243-253.

Scott, C. (1972). On the evaluation of verbal material in parapsychology: A discussion of Dr. Pratt's monograph. *Journal of the Society for Psychical Research, 46*, 79-90.

Solfvin, G. F., Kelly, E. F., Burdick, D. S. (1978). Some new methods of preferential-ranking data.. *Journal of the American Society for Psychical Research, 72*, 93-109.

Stahl, S. (2004). Permutation-based methods for examining confusion data in ESP experiments. *Journal of Parapsychology, 66*, 381-403.

Storm, L. (2003). Remote viewing by committee: RV using a multiple agent / multiple percipient design. *Journal of Parapsychology, 65*, 325-342.

Stuart, C. E. (1942). An ESP test with a drawing. *Journal of Parapsychology, 6*, 20-43.

Targ, R. (1994). Remote-viewing replication: Evaluated by concept analysis. *Journal of Parapsychology, 58*, 271-284.

Utts, J. M. (1989). Randomness and randomization tests: A reply to Gilmore. *Journal of Parapsychology, 53*, 345-351.

Utts, J. M. (1993). Analyzing free response data: A progress report, in L. Coly & D. S. McMahon (Eds.) *Psi research methodology: A re-examination.* New York: Parapsychology Foundation, pp. 71-83.

Wald, A., & Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *Annals of Mathematical Statistics, 15*, 358-372.

Watt, C., & Nagtegaal, M. (2004). Reporting of blind methods: An interdisciplinary survey. *Journal of the Society for Psychical Research, 68*, 105-114.

Welch, W. J., & Gutierrez, L. G. (1988). Robust permutation tests for matched-pairs designs. *Journal of the American Statistical Association, 83*, 405-455.

Appendix

Description of the building process

```
{Determine the number of objects (n), evaluate the actual
statistic and store in a variable (Observed), initialize a
counter as one (Index), initialize a pointer to the location
in the permutation (Position), and initialize an "empty"
n x 1 array as zeros (permutation).}

 while (Position ~= 1) || (permutation(Position) <= n)
    Position = Position + 1;      Forward Branch
    fathom = 0;
    while fathom == 0
        {Right Branch by loading the next available object,
        into permutation(Position).  If/when
          permutation(Position) > n, all objects have been
          considered.}
          while ((Redundancy == 1) &&
          (permutation(Position) <= n))
                permutation(Position) =
                permutation(Position) + 1;
                Redundancy = 0;
                 for k = 1:(Position - 1)
                     if permutation(Position) == permutation(k)
                         Redundancy = 1;
                     end
                  end
          if (Position == 1) && (permutation(Position) > n)
          Termination break;
        end
        if (Position > 1) && (permutation(Position) > n)
        Retraction permutation(Position) = 0;
            Position = Position - 1;
        else
              if Position == n      Complete Sequence is
              Ready for Evaluation
                 {Evaluate the statistic for the current
                 permutation as a variable, stat.}
                  if stat >= Observed
                      Index = Index + 1;
                  end;
              else
                  fathom = 1;
              end
          end
        end
    end
```

```
{Return p = (Index/factorial(n)) as the p-value.}
```

The bounding evaluation is inserted after the "fathom = 1" command:

```
PREstat = 0; LowerBound = 0; UpperBound = 0;
for i = 1:n
    found = false;
    for j = 1:Position
        if permutation(j) == i
            found = true;
            PREstat = PREstat + A(j, permutation(j));
        end
    end
    if ~found
        largest = A(Position + 1, i);
        smallest = A(Position + 1, i);
        for k = Position + 1:n
            if largest < A(k, i)
                largest = A(k, i);
            end
            if smallest > A(k, i)
                smallest = A(k, i);
            end
        end
        LowerBound = LowerBound + smallest;
        UpperBound = UpperBound + largest;
    end
end
fathom = 1;
if PREstat + LowerBound >= Observed
    fathom = 0
    Index = Index + factorial(n - Position);
else
    if PREstat + UpperBound < Observed
        fathom = 0;
    end;
end
```

An Experiment with Covert Ganzfeld Telepathy

Peter Pütz, Matthias Gäßler, & Jiří Wackermann

Institute for Frontier Areas of Psychology and Mental Health
Department of Empirical and Analytical Psychophysics
Freiburg i.Br., Germany

Abstract

The aim of our study was to test a modified ganzfeld telephathy procedure, which conceals the intended anomalous information transfer. Forty pairs were recruited for ganzfeld sessions, each comprising three trials consisting of a 'communication' and a 'rating/reporting' phase. During the 'communication' phase (20 min), one member of the pair (A) was exposed to multimodal ganzfeld and reported her/his imagery, while the other (B) memorised a repeatedly presented video clip. In the 'rating/reporting' phase subject A rated the similarity of the 'target clip' and three 'decoys' to the ganzfeld imagery, while simultaneously subject B gave a written account of the content of the presented target. Trials in which the highest score was assigned to the target clip were considered as correct identifications. In 39 out of 120 trials (32.5%) the presented target clip was correctly identified ($p = .039$). Statistics based on ranks of all four video clips revealed no significant deviations from chance expectancy. The modified experimental procedure (a) yields correct identification rates comparable with the traditional procedure, (b) allows study of 'ganzfeld telepathy' without confronting subjects with an 'impossible task'.

Introduction

Dyadic communication in the ganzfeld ('ganzfeld telepathy') is an established paradigm in experimental parapsychology for the last few

---

decades. The results of these experiments have often been considered as experimental evidence for an anomalous information transfer in the ganzfeld (Honorton, Berger, Varvoglis, Quant, Derr, Schechter & Ferrari, 1990; Bem & Honorton, 1994), although this conclusion was questioned by later meta-analyses (Milton & Wiseman, 1999).

Since the beginning of ganzfeld telepathy experiments in the early 1970s, the procedure has been repeatedly modified. These changes include the invention of an automated ganzfeld procedure ('auto-ganzfeld') through Berger and Honorton (1985) to overcome shortcomings related to the early ganzfeld experiments, in which manual target randomisation and ratings recording etc. were used, and use of dynamic targets (video clips). Some later studies based on an improved version of the automated ganzfeld procedure ('digital autoganzfeld') used multiple trials per session to increase the statistical power and to help identifying pairs able to establish anomalous communication. Four trials per session ('serial ganzfeld') were used by Parker and Westerlund (1998); Goulding, Westerlund, Parker and Wackermann (2004) used two trials per session.

Except for these modifications, ganzfeld telepathy experiments have several key elements in common which are of various, often unclear or disputable importance:

(a) Participants are usually fully aware of the intended anomalous information transfer. The subject in the ganzfeld attempts to 'receive' the video clip (or other target material) his/her partner watches, and the latter intends to 'transmit' the content of the video clip. Therefore, participants with a preconceived interest or belief in the 'paranormal' will perceive the task differently from participants without such beliefs, who may find the task strange, ridiculous, or may be facing a 'mission impossible' situation.

(b) The 'receiver' is allowed and/or encouraged to continuously verbalize her/his mentation, which is recorded for later evaluation. However, continuous verbalisation may be problematic for the following reasons. Firstly, the rationale for using ganzfeld in parapsychological research was to induce the allegedly psi-favourable 'internal attentional state' (Honorton, 1978). Indeed, prolonged exposure to ganzfeld stimulation frequently induces dream-like, pseudo-hallucinatory imagery. However, the continuous verbalisation may contaminate the genuine ganzfeld imagery, and it may even counter-act it altogether, diverting the subject towards thought fragments, free associations, and

other cognitive processes. Secondly, if one wants to combine ganzfeld experiments with simultaneous measurements of the brain's electrical activity, continuous verbalisation would inevitably cause contamination of the EEG data with muscular artifacts.

(c) It has often been argued that targets of rich, variable, emotional content and dynamic character facilitate the 'psi' communication (Bem & Honorton, 1994; Parker, Grams, & Petterson, 1998); 'good targets' should be meaningful and have human interest (Watt, 1988). However, regarding the dynamic character, we should note that static targets (photographs or drawings) were used, reportedly with success, in early ganzfeld studies (Honorton, 1985a, 1985b). As to the content variability issue, we are facing contradictory claims: for example, in remote viewing research rather homogeneous stimulus material is preferred, which is based on the rationale that more homogenous stimuli lead to 'noise reduction' (May, Spottiswoode & James, 1994). Lantz, Luke and May (1994) reported a significant difference between static and dynamic targets, favouring static targets in a telepathy experiment. In another telepathy experiment (without sender), topically restrictive dynamic targets showed a significant increase of anomalous cognition compared to the unbounded dynamic target pool used in the previous experiment. We should add that the use of heterogeneous stimuli makes post hoc analyses of possible relations between stimulus content and anomalous information transfer rather difficult.

The aim of this explorative study was elaboration of an experimental protocol stripped down of most traditionally employed components. We opted for a 'minimalised' procedure which did:

(i) not disclose the intended anomalous information transfer and would be thus acceptable for all participants ('non-overt telepathy'); (ii) focus on the ganzfeld-specific imagery, avoid continuous verbalisation but allow comprehensive reporting of ganzfeld-induced experience; (iii) use sets of stimuli with maximal within-set content diversity, constructed from homogenous stimulus material, and based on an objective measure of stimulus content differences.

Of main interest was the performance of the participants in terms of target identification. All other reported statistics were post-hoc analyses.

Methods

Participants

Forty pairs (48 female, 32 male; mean age: 25.7 years, range: 16.8–55.3 years) were recruited for the experiment via the local university's job exchange service and a newspaper advertisement.

With one exception all participants were reportedly of good health and had no medical or neurological problems. The exception was a female participant who was subject to anticonvulsive medication against idiopathic grand mal seizures, but seizure free for the last two years. As the experiment did not involve EEG recordings, the pair was not excluded from the sample. One of the examined pairs were female twins (age: 22 years).[1]

The participants were not aware of the aim of the study, i. e. anomalous information transfer in the ganzfeld; the study was described in the advertisements as 'an experiment in perception and relaxation'. Before the experiment, the participants signed a written consent not to reveal the information about the study to a third party; they were informed about the proper intent of the study only after the experimental session.

Questionnaires and inquiries

A standard participant information form (PIF) was used to collect the subjects' sociodemographic data, their general mental and physical status and their medical history. A short status questionnaire was applied to assess their physical and mental condition immediately before the experiment.

To assess personality traits of the participants we used the *NEO Five Factor Personality Inventory* (Costa & McCrae, 1992) in a German translation by Borkenau & Ostendorf (1993), a questionaire which we also used in our earlier ganzfeld studies.

The relationship between the participants was assessed by a special response form: the duration of the relationship (years/month), the kind of relationship (acquaintance, friends, intimate friends, partner, spouse) and its intensity. The latter was measured by placing a mark on a 100 mm preprinted line segment, with endpoints labelled 0 (unknown) and 100 (maximum), and an anchor point at 10 mm = a person known from seeing, no acquaintance.

---

[1]This pair participated in two sessions, but only the results of the first session were included into the data of the present study; details are given in the appendix.

Figure 1. Example of a set of four video clips. Representative frames of the respective video sequences are shown: (a) Preparation of a meal (casserole) [content code: HumArtNat]; (b) train of crawling caterpillars [content code: AniEle(earth)]; (c) cathedral in the Normandy [content code: HumArc]; (d) burning Christmas tree [content code: ArcArtEle(fire)]. For detailed explanation of the content codes see text.

During the experimental sessions (see the experimental procedures section below) a shortened version of an inquiry developed in our laboratory (Pütz et al., 2006) was used, assessing sensory modalities involved in reported percepts, distinctness and vividness, and various other aspects of the reported ganzfeld imagery.

Stimulus material

A database of 82 video clips was collected from publicly available sources (Internet, video tapes libraries etc.), using the following selection criteria: (a) understandable and (prima faciae) interesting content, (b) content homogeneity, and (c) minimal duration 30 seconds. Clips fulfilling the above-given criteria were mostly taken from documentary movies.

The next step was grouping of selected video clips to groups of four ('4-sets') to be used in the experiments (in each trial, one clip served as the 'target' stimulus and the three remaining clips as 'decoys'). A total of eight 4-sets were selected from the primary database, using the procedure described below (see, for example, Figure 1).

A stimulus content classification system (SCCS) was developed for the purpose of the present study and used to classify the contents of the database. The SCCS has six primary content categories: humans (Hum), animals (Ani), architecture (Arc), human-made objects or artefacts (Art), nature or natural sceneries (Nat) and 'elements of Nature' (Ele), i. e. fire, water, air, and earth.

Each clip is thus described by a 6-dimensional binary vector, $\mathbf{x} = (x_1, \ldots, x_6) \in \mathbb{B}^6$, where $\mathbb{B} \equiv \{0, 1\}$; 1 encodes presence and 0 encodes absence of the respective content category. The space of all possible combinations, $\mathbb{B}^6$, thus consists of $2^6 = 64$ elements. Contents difference between two video clips was measured by the Hamming distance,[2]

$$d_H(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{6} |x_i - y_i|; \quad \mathbf{x}, \mathbf{y} \in \mathbb{B}^6.$$

Contents diversity of a 4-set, $D(S)$, is defined as the sum of all possible pair-wise Hamming distances within the set,

$$D(S) = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{y} \in S} d_H(\mathbf{x}, \mathbf{y})$$

(the maximal possible diversity per 4-set is 24).

An iterative optimisation procedure was used to generate 4-sets from the available database with (a) maximised contents diversity for each 4-set, (b) yielding as many 4-sets as possible. The database allowed a maximum of only six 4-sets reaching the maximal contents diversity, $D(S) = 24$. The aim was to maximise the overall contents diversity,

$$\mathcal{D} = \sum_{j=1}^{N} D(S_j),$$

while obtaining a sufficiently large pool of 4-sets. Balancing the pool size and diversity, the procedure resulted in eight 4-sets deviating only by 4.7% from the theoretical maximum of the overall contents diversity $\mathcal{D}$.

As shown in Table 1, the selection procedure compensates the non-uniformity of relative occurrences of the SCCS-categories in the available database, approximating the theoretical value 0.5 which would be

---

[2] Hamming distance (Hamming, 1950) is a standard tool in information coding and transmission theory, but also widely used in diverse areas of science and engineering as cryptography, pattern recognition, image analysis, and analysis of genomic sequences (He, Petoukhov, & Ricci, 2004).

Table 1: Relative frequencies of SCCS-categories

	Hum	Ani	Arc	Art	Nat	Ele
Entire database	0.39	0.30	0.40	0.25	0.40	0.44
Selected clips	0.50	0.38	0.50	0.44	0.44	0.50

Table 2: Descriptive data of the stimulus material

	Entire database	Selected clips
Minimal duration (seconds)	30.0	31.0
Maximal duration (seconds)	172.0	172.0
Mean duration (seconds)	68.5	69.1
Mean number of categories/clip	2.17	2.75

achieved for all categories by a complete and uniform coverage of $\mathbb{B}^6$. Table 2 contains descriptive data of the video clip database and of the 32 selected video clips.

Apparatus

Ganzfeld stimulation, room A: The same procedure for the multi-modal ganzfeld (MMGF) was used as in our earlier studies (Wackermann, Pütz, Büchi, Strauch & Lehmann, 2002; Pütz et al., 2006): The subjects' eyes were covered with semi-translucent goggles (anatomically shaped halves of ping-pong balls) and illuminated with a red-coloured 60 Watt incandescent lamp, from a distance of ~120 cm. Monotonous sound of a waterfall was played back via headphones.

The room *A* was equipped with a 'voice-key', which was triggered by the onsets of subjects' imagery reports; the device generated digital signals, which were transmitted to a computer in the adjacent room where they were stored.

Video presentation, room B: A modified version of the 'Automated Digital Ganzfeld' software, developed at the University Gothenburg (Goulding et al., 2004) based on a MS Windows Media-Player plug-in (Version 6.4), was used for the presentation of the stimulus material and recording of the imagery reports. The software transmitted digital signals marking beginning and end of each trial, and beginnings of the repeated target clip presentations, to the computer in the adjacent room, where they were stored in parallel with the report onset markers.

The video clips were presented on a 17" XGA Acer FP752 TFT display, at native resolution 1024×768 and at the standard monitor refresh

rate of 60 Hz, thereby applying a frame rate conversion 50 Hz to 60 Hz for clips based on PAL sources. All video clips were encoded in MPEG-2; 29 in PAL format (720 × 576 pixel), two in NTSC format (720 × 480 pixel), and one video clip was of resolution 382 × 280. The mean effective bit rate of the video clips was 4450 kbps. The distance of the participants to the TFT display was ≈ 75 cm, the angle of vision of the presented video clips was $20°$ horizontal and $17.6°$ vertical.

Experimental procedures

After the participants were introduced to the laboratory and the two experimenters, they filled in the questionnaires. They were then separated and obtained individual detailed instructions according to their assigned task in the experiment. As a rule, the participant who completed the questionnaires earlier was assigned to the ganzfeld. In the following the two subjects are named A = the subject exposed to the ganzfeld, and B = the subject watching the video clip;[3] during the experimental session they were accompanied by two experimenters, referred to as E_A and E_B, respectively.

Subject A was introduced to the laboratory room A and explained the ganzfeld procedure. (S)he was instructed to report ganzfeld imagery, if it occurred, at the moment it was maximally pronounced or just about to vanish. Subject B was introduced to the laboratory room B and instructed to watch a short video clip presented (without sound) on the display, and to memorise its contents for a latter recall and written report. Each pair served in one experimental session, which comprised three trials; each trial consisted of a 'communication' phase, followed by a 'rating/reporting' phase.

Communication phase: During this phase participant A was exposed for 20 min to multi-modal ganzfeld (MMGF) in room A, while participant B watched a target video clip in room B. At the onset of the subject A's report, experimenter E_A stopped the acoustical stimulation and the subject gave a free verbal account of the imagery; afterwards, (s)he answered the ganzfeld inquiry (see the questionnaires and inquiries section above). The MMGF stimulation was then continued until the next verbal report, or the end of the 'communication phase'.

---

[3] In the usual jargon, A and B thus refer to the 'receiver' and 'sender', respectively. For reasons that will become obvious from the following description of the procedure, we abstain from the traditional nomenclature.
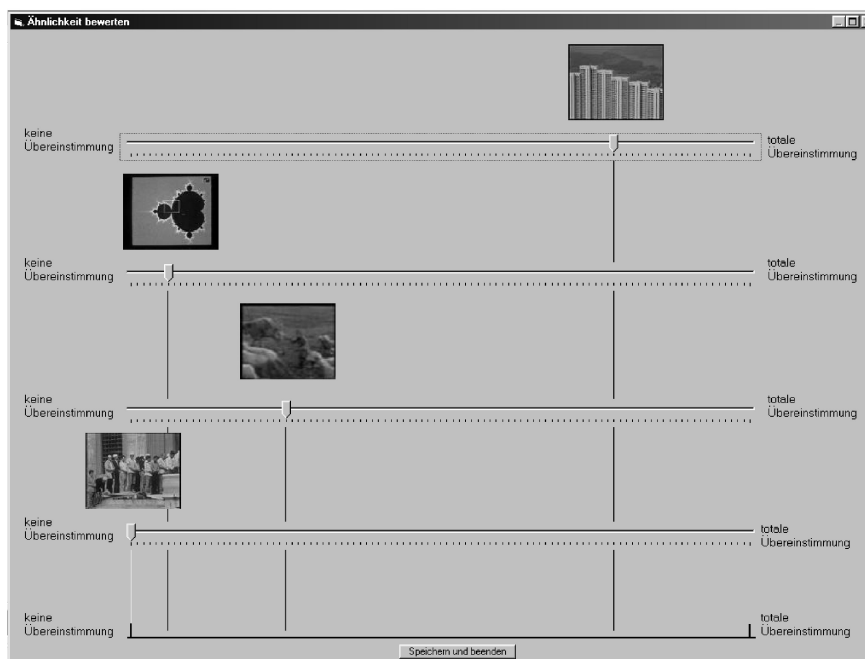
Figure 2. Screenshot of the evaluation software, used for the rating of similarity of the ganzfeld induced experience to video clips.

The subject B's task was to watch a target video clip presented repeatedly in room B, for a total duration of 20 min. Each iteration started immediately after the end of the previous presentation. The subjects were encouraged to follow the presentation as attentively as possible; however, they were also allowed to close their eyes and exert a 'mental replay' of the clip, to avoid fatigue or 'overload'. As soon as the 20 min presentation-loop was over, participant B was guided to room C for the recall of the watched video clip. Afterwards, the ganzfeld stimulation was stopped and subject A guided into room B for the rating.

Rating/Reporting phase: In room B subject A was presented four video clips (the target clip and three 'decoys' from the same 4-set) in random order, and asked to rate the degree of similarity of each of the four clips to his/her prior ganzfeld experience. The ratings were assigned by positioning mouse-operated 'sliders' on a scale ranging from 'no similarity' (0) to 'maximal similarity' (100) (Figure 2).[4] The subject could freely choose the sequence in which (s)he watched the clips and made her/his rating. Simultaneously, subject B (room C) gave a written account of the watched video clip, using a form of her/his choice:

---

[4]We should note that the evaluation software used for the ratings does not allow to award the same rating to several video clips, to ascertain unequivocal assignment of ranks.
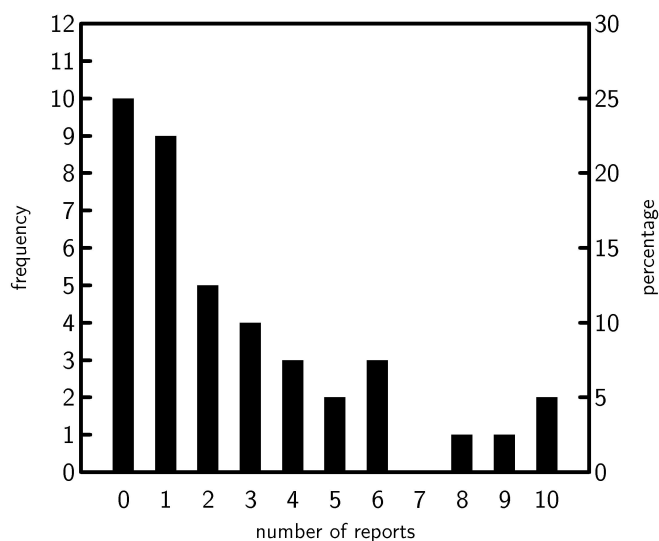
Figure 3. Distribution of report frequencies per session.

written word, drawn pictures, or any combination thereof.

As soon as subject A finished the rating, (s)he was guided back to room A and prepared for the next trial. Experimenter E_A gave a signal to experimenter E_B, who stayed meanwhile in another room, D, via a phone call. Afterwards experimenter E_B told subject B that the time for reporting was over, and accompanied her/him back to room B, where the next trial was initiated.

The 40 experimental sessions were acquired in four blocks of ten sessions each, within a time-span of 13 months. The time periods needed to accomplish one 10-session block varied from 16 to 56 days (mean = 38 days).

Results

A total of 108 imagery reports were collected, that is, in the mean average, 2.7 reports per session. The average yield of the first, second and third trial in a session was 1.0, 0.8 and 0.9, respectively. Figure 3 shows the distribution of imagery report frequency. The U-shaped bi-modal distribution suggests large inter-individual differences in responsiveness to the MMGF. Roughly summarised, about 60% of participants gave less-than-average number of reports, in contrast to a small group of 'high responders' (≥ 8 reports/session).

Imagery reports

Relative frequencies of involved sensory modalities were comparable to those reported in earlier studies (Table 3). Ganzfeld imagery was predominately of visual nature, acoustic imagery being the second most frequent sensory modality: relative frequency of other sensory modalities was rather low.

Table 3: Incidence of reported sensory modalities in ganzfeld imagery experiments: SOGF = comparison of sleep onset and ganzfeld imagery (Wackermann et al., 2002), GFS/GFE = screening for 'high-responders' and data from selected 'high-responders' (Pütz et al., 2006), ADGF = data from the reported study. Shown are relative frequencies in percentages. Note that the columns sums >100%, indicating that some of the imagery episodes involved more than one modality.

Modality	SOGF	GFS	GFE	ADGF
Visual	90.4	94.3	97.6	85.2
Acoustic	28.8	16.1	23.2	24.1
Olfactory	16.4	3.2	3.7	0.0
Tactile	26.0	9.7	8.5	9.3
Kinaesthetic	0.0	5.4	2.4	7.5

Similarity ratings

The data collected in the rating phase (see the methods section) consists of 40 (pairs) × 3 (trials) = 120 data vectors. Each of these vectors contains four similarity ratings (0–100 scale) of the four video clips in the given 4-set. For the purpose of further analyses, these ratings were sorted in a descending order and transformed into ranks; that is, rank '1' corresponds to the highest score, and rank '4' to the lowest score.

Of particular interest are cases when the highest rating was assigned to the video clip actually presented to subject B ('target'). If the subject A's task were solely to indicate the clip of the highest degree of similarity (forced choice), these cases would correspond to 'direct hits' in the usual nomenclature of ganzfeld telepathy experiments. Therefore, the cases in which the target clip was given rank '1' (highest rating) are in the following referred to as *Correct Target Identification* (CTI).

Statistics of ranks

By single trials: The null hypothesis H_0 predicts a uniform distribution of the ranks '1–4', with probabilities .25. Observed frequencies do not deviate significantly from the theoretical distribution (see Table 4); $\chi^2 = 4.400$; *df* = 3; p = .221. The distribution of the observed values
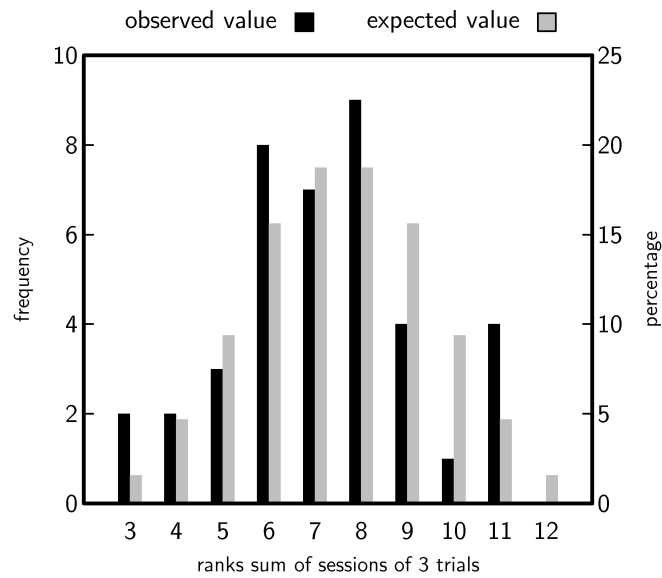
Figure 4. Theoretical and observed distribution of the sum of ranks of the three presented target clips in one session.

suggests that mainly relative frequencies of rank '1' and '2' differ from chance expectancy.

Table 4: Distribution of ranks assigned to target video clips

Rank	Count	Relative frequency
1	39	0.325
2	23	0.192
3	29	0.242
4	29	0.242

By sessions: For each session we take the sum of ranks of the presented target clip, from trials 1–3. The theoretical distribution of these rank sums, predicted by the H_0 ranges from 3–12 (mean = 7.5) and is easily obtained by complete enumeration. The theoretical distribution, and the observed rank sums, are shown in Figure 4. The mean observed rank sums is 7.2, which is not significantly different from the mean chance expectation, 7.5. Noteworthy is a marked asymmetry of the observed distribution, with the obviously deviating values for rank sums 3 and 12.
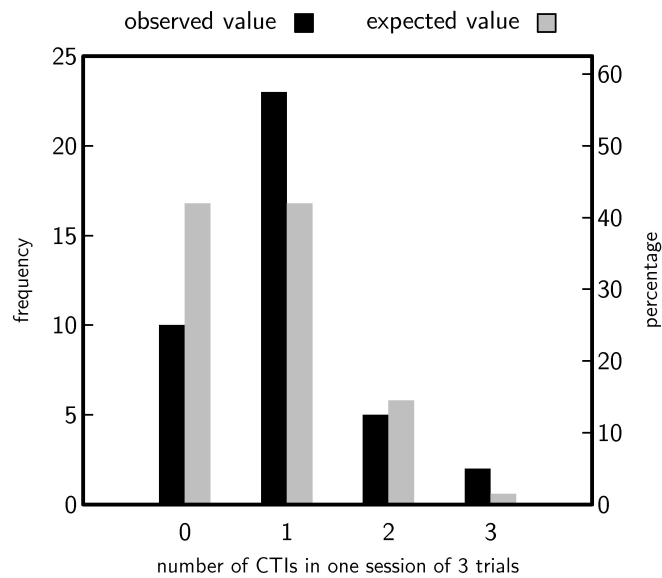
Figure 5. Distribution of correct target identifications per session. Theoretical distribution is shown in gray colour, observed frequencies are shown in black.

Statistics of correct target identifications

By single trials: Here we focus on the total number of correct target identifications. If the subjects were assigning their ratings by chance (H_0), we should expect 30 out of 120 trials (25%) to be correctly identified. The observed number of CTI = 39 corresponds to a 'hit-rate' of 32.5%, which is significantly higher than the mean chance expectation ($p = .039$, binomial distribution $\mathbf{B}_{120}(.25)$).[5]

By sessions: The above-reported p-value for the deviation of the CTI rate from the MCE implies a Bernoullian model of $N = 120$ independent trials with a constant probability of success, .25. This, however, is not quite an adequate model for the given experimental design, as the CTIs resulted from three repeated trials for each pair/session. Hence, it is more appropriate to treat the outcome of each session as an independent data unit (similarly as we have studied sums of ranks in the preceding section). The total of CTIs per session can attain values from 0 through 3. If, for the subject A in a given session, the probability of the CTI is .25 (as predicted by H_0), the probabilities of obtaining 0,1,2, or 3 CTIs are determined by the binomial distribution $\mathbf{B}_3(.25)$; this evaluates to

---

[5] Here and in the following, $\mathbf{B}_n(p)$ denotes the binomial distribution of successful outcomes from n trials, with success probability p.

$p_0 = p_1 = .4219$, $p_2 = .1406$, $p_3 = .0156$. Figure 5 shows the theoretical distribution (H_0) and the observed distribution of CTI/session.

A comparison of the observed and theoretical distribution, based on the 'classic' Pearson's χ^2 statistics, yields $\chi^2 = .119$, *df* = 3, $p = .044$. However, this result is not trustworthy because of extremely low frequencies in one of the categories (3 CTI/session). Therefore, we should prefer the 2I-test, which is designed for the same purpose but more robust (Weber, 1980, p. 194ff). The 2I-statistics is 7.254, that is, below the critical value for 3 *df* ($p = .064$); hence we consider the result as merely suggesting a better-than-chance performance in target identifications.

Extreme performance

Two subjects correctly identified all three targets in one session ('hat-trick'). As shown above, the probability of a 'hat-trick' response is $p_3 = .0156$; thus the two 'hat-tricks', taken as singular events, suggest at the first sight a 'significant' result. However, the binomial probability $\mathbf{B}_{40}(p_3)$ to get *at least* two 'hat-tricks' in a series of 40 sessions evaluates to $p = .129$, indicating that the occurrence of 'hat-tricks' is not much of a surprise. Incidentally, one of the two pairs producing three CTIs in a session were twins (this pair also participated in another session, not included in statistical evaluation; see the appendix).
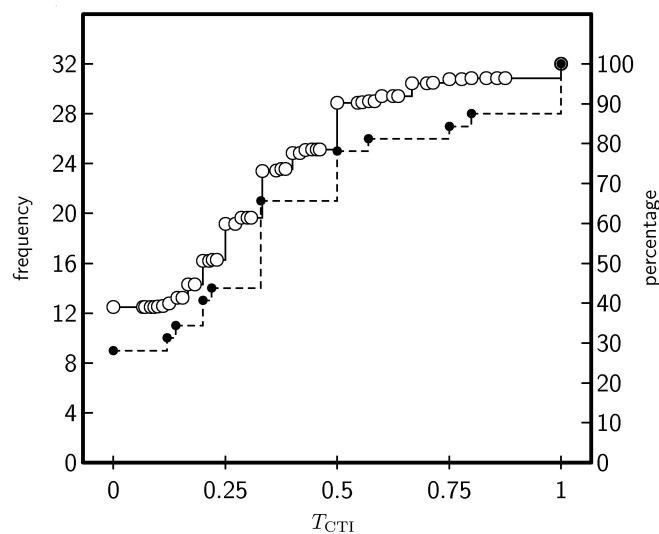

Figure 6. Cumulative frequency distribution of target-specific identification rates (T_{CTI}) observed in our study (dots), plotted against a theoretical distribution (open circles) estimated via a Monte-Carlo simulation (see text).

Table 5: Correlation between NEO-FFI personality factors and CTI/session. Row A: participants exposed to MMGF; row B: participants watching the target clip. Shown are Spearman correlation coefficients, values in bold font are statistically significant ($p < .01$).

NEO-FFI	N	E	O	A	C
A	.125	.166	−.290	.196	**.438**
B	−.230	−.082	−.041	.054	.152

Correlations between CTIs and other variables

Spearman's rank correlation coefficients between numbers of CTI/session individual variables such as ganzfeld productivity (e. g. number of reports per session), personality factors, interpersonal relationship, and physical/mental status of the participants, were calculated for both groups A and B.

Imagery productivity: Correlation between the number of CTI/session and the number of imagery reports per session (group A) was almost exactly zero ($r = .036$, $df = 38$, $p = .825$), thus indicating no relationship.

Interpersonal relationship: The only noteworthy correlation between interpersonal relationship intensity (group B) and CTI ($r = −.29$, $p = .09$) is not statistically significant.

Personality factors: Five personality factors, Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A) and Conscientiousness (C), were assessed by means of the NEO-FFI. Table 5 shows correlations between these personality factors and CTI/session for both groups of participants, A and B. The only significant, and remarkably high, correlation was found for the personality factor Conscientiousness in participants A ($r = .44$, $p = .005$).

Status variables: For participants A two variables from the status inquiry before the experiment were significantly or almost significantly correlated with the number of CTI/session: 'alertness' ($r = .29$, $p = .07$) and 'emotional condition' ($r = .31$, $p = .05$).

Target-specific identification rates

The obviously non-uniform distribution of ratings across the video-clips suggest that some clips were 'favoured' by the participants, that is to say, they were frequently given the highest similarity score whenever the respective 4-set was used. For example, the 4-set shown in Figure 1 was used nine times in the entire study; the 'caterpillar clip' (Figure 1b) was used four times as the 'target', and in all four instances given the highest score, i.e., 'correctly identified'.

The experimental procedure principally allows re-use of stimulus material (similarly to the 'open deck' strategy) and thus such non-uniform re-occurrences of the same set/stimulus are to be expected. This, however, lets the question arise whether the observed non-uniformity of CTIs across clips is caused by an unknown factor — perhaps 'anomalous cognition'? — or are rather due to the fact that some video clips are more 'appealing' to the subjects than others (a sort of 'stacking effect'). It is thus of interest to see if certain stimulus contents are better suited for anomalous information transfer. For this purpose, we examine 'target-specific identifications rates', defined as:

$$T_{\mathrm{CTI}} = N_{\mathrm{CTI}}/N_{\mathrm{shown}}$$

where N_{CTI} is the number of times a target video clip was correctly identified and N_{shown} the number of times the video clip was used as a target.

Figure 6, showing the cumulative frequency distribution of T_{CTI} in our study demonstrates a relative deficit of target clips that were *never*
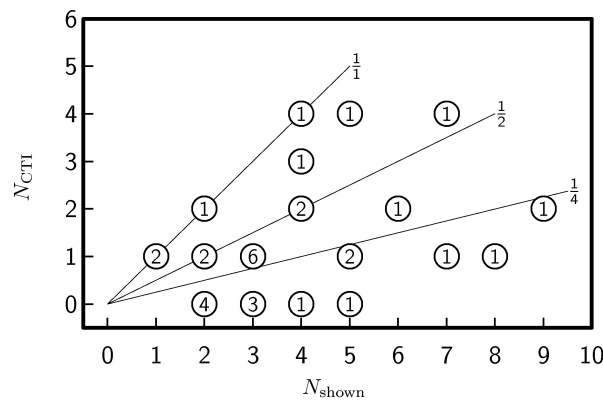

Figure 7. Plot of the numbers of correct identifications (N_{CTI}) versus the numbers of target usage (N_{shown}). Three target-specific identification rates are shown for reference: T_{CTI} = .25 (mean chance expectation), .5, and 1. Figures in circles denote target clip occurrences.
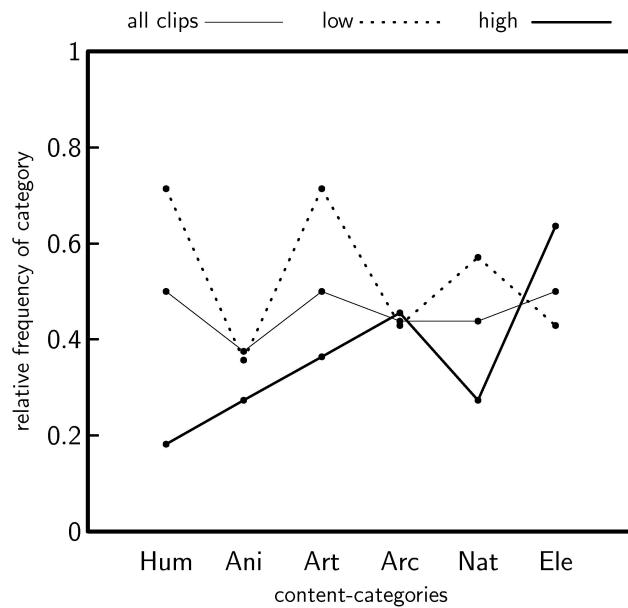
Figure 8. Relative frequencies of content categories shown separately for the two subsets defined by extreme T_{CTI}'s, S_{low} and S_{high}, and the entire pool. Abbreviations: Hum = Humans, Ani = Animals, Art = Artefacts, Nat = Nature, Ele = Elements.

identified ($T_{\mathrm{CTI}} = 0$). By contrast, four out of 32 video clips were *always* correctly identified ($T_{\mathrm{CTI}} = 1$). The content codes of these four clips were: 'animals' and 'elements', 'humans' and 'architecture', 'architecture' and 'elements', and 'animals'. The first two of the four just mentioned clips belonged to the same 4-set. To estimate the probability to get *at least* four target clips with $T_{\mathrm{CTI}} = 1$ in an experiment of given design (120 trials, 32 clips), a Monte-Carlo simulation of 10000 such experiments was carried out, yielding $p = .023$.

A plot of the numbers of correct identifications versus the numbers of target usage (Figure 7) reveals that those four video clips with $T_{\mathrm{CTI}} = 1$ account for eight of the total 39 correct target identifications, that is, 20.5%. Eleven targets have $T_{\mathrm{CTI}} \geq .5$ and account for 64.1% of all CTIs; in other words, almost 2/3 of observed CTIs are based on only 1/3 of the stimulus material.

To examine differences in the content categories for targets with high and low identification rates, two subsets were drawn from the stimulus material, based on a median-split at .333: S_{low} consisting of clips with $T_{\mathrm{CTI}} < .333$ ($n = 14$), and S_{high} consisting of clips with $T_{\mathrm{CTI}} > .333$ ($n = 11$). A visual inspection of the 'content profiles', i. e. of relative frequencies of the six SCCS categories, for the two sets, S_{low} ver-

sus S_{high}, reveals that targets with high T_{CTI} were generally 'simpler' in terms of the contents than those with low T_{CTI} (Figure 8). Mean numbers of content categories for subsets S_{low} and S_{high}, were 3.35 and 2.09 respectively. This applies to five of six content categories, with an exception of the category 'elements' which is more frequently present in S_{high} targets (63.3%) than in S_{low} (42.9%).

Discussion and conclusion

The modified experimental procedure yields 'hit-rates' comparable to figures reported from traditional ganzfeld-telepathy experiments, even if the participants had no intent to establish a 'telepathic communication' and, in fact, were not selected for their belief in the possibility of such communication.

Average yield of imagery reports was lower than in our earlier study (Pütz et al., 2006); this, however, was expected, as the subjects were not pre-selected for 'ganzfeld responsiveness' and none of them had former experience with ganzfeld. Given the lack of a correlation between CTI/session and imagery productivity, it is quite possible that the genuine ganzfeld imagery is not directly related to, or necessary for, anomalous cognition.

The observed rate of correct target identifications, 32.5%, is significantly higher than the mean chance expectation. However, statistics based on CTI/session only approached the conventional limit of 'significance', and statistics based on ratings of all four video clips in a respective set did not show a significant deviation from H_0. Therefore it would be premature to interpret the results as indicative of an anomalous information transfer. We still cannot fully rule out the possibility of a 'stacking effect' (see above) or other, unknown sources of the observed effect.

Noteworthy, Goulding et al. (2004) obtained results close to chance level, using basically the same software as in the present study but different stimulus material. In their study, the choice of video clips was based on rather subjective criteria: "the clips chosen were clips that [the experimenters] thought would be interesting and meaningful for the participants." (Goulding et al., 2004, p. 79). By contrast, the selection of the stimulus material for our study was based on pre-defined, content-related criteria (see the section on stimulus material for further details).

This leads to the problem of the choice of suitable stimulus mate-

rial for ganzfeld-telepathy experiments, or experiments in anomalous cognition in general. Our results seem, on first sight, to be rather in line with the findings of May et al. (1994) and Lantz et al. (1994) who preferred homogeneous stimuli in remote viewing experiments. In our study, the comparison of targets with low and high T_{CTI} suggests that homogeneity and/or 'topical restriction' are related to higher identifications rates. Here the notion of homogeneity applies to *single stimuli*; however, maximal content diversity is arguably required on the level of *stimulus sets*. Therefore, the entire sets should be as 'rich' as possible, in other words, heterogeneous in terms of within-set content differences. For this purpose we used the above-described content-classification and set-construction procedures. This is, to our knowledge, the first ganzfeld-telepathy study where such strictly formalised criteria have been applied.

It is also worth mentioning that stimuli with high T_{CTI}s often included the 'elements', i. e. water, air, earth or fire. We may assume that such stimuli, often of amorphous appearance, may remind of the ganzfeld exposure — indeed, ganzfeld is often described by participants as a 'diffuse red mist or fog'. This would be a trivial explanation for increased similarity ratings (and thus for increased frequency of rank '1' scores), but would not *per se* explain the increased *correct identification rates* (unless these are due to a 'stacking effect'). Or is perhaps the 'elements' category better suited for anomalous information transfer?

Given that many open questions as to the nature of the observed effect, our interpretation of correlations between the CTI performance and personality or other individual factors can be only tentative. Interestingly, it was only the personality factor 'Conscientiousness,' i. e. determination and goal orientation, which was positively correlated with CTI rates per session, while 'Extraversion', a personality trait frequently connected to success in 'psi tasks' (Honorton, Ferrari & Bem, 1992) was not correlated to the CTI performance. Further, we did not find any relation between interpersonal relationship and CTI performance. The correlations with status variables suggest that participants who were more alert and in higher mood at the beginning of the experiment were performing better in terms of CTI. According to these results, participants who were more focused and compliant with the experimental situation were more successful in target identification than those with a less compliant attitude.

Our findings question the alleged importance of the participants'

attitudes and beliefs concerning anomalous cognition or 'psi', or even of their being aware of the aim of the experiment. Thus it seems unnecessary to insist on 'belief in psi' as a selection criterion. Using the cover story, the proper aim of the experiment is concealed from the participants; there is no risk that 'skeptical' participants would be facing an 'impossible task'. Consequently, the modified experimental procedure allows to study dyadic communication in ganzfeld with general population, or samples selected by other criteria unrelated to anomalous cognition. An important component of the procedure is selection/construction of the stimulus material, using formalised, objective criteria. Last but not least, the method used to collect reports of ganzfeld-induced subjective experience is compatible with simultaneous electrophysiological recordings.

Finally, we would like to quote from Bem, Broughton & Palmer (2001, p. 215), who argued that "[p]erhaps there is some merit in continuing to conduct exact replications of the ganzfeld procedure, but genuine progress in understanding psi rests on investigators' being willing to risk replication failures by modifying the procedure in any way that seems best suited for exploring new domains or answering new questions." We feel that our study suits well this programmatic thesis.

Acknowledgements

References

Bem, D. Broughton, R., & Palmer J. (2001). Updating the ganzfeld database: A Victim of its own success? *Journal of Parapsychology*, *65*, 207–218.

Bem, D. & Honorton, C. (1994). Does psi exist? Evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*, 4–18.

Berger, R. E. & Honorton C. (1985). An automated psi ganzfeld testing system. In *The Parapsychological Association 28th Annual Convention: Proceedings of Presented Papers Volume 1*, (pp. 1–36). New York, NY: Parapsychological Association.

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) Handanweisung.* Göttingen: Hogrefe Verlag für Psychologie.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory. Professional Manual.* Odessa, Florida: Psychological Assessment Resources.

Duane, T. D., & Behrendt, R. (1965). Extrasensory electroencephalographic induction between identical twins. *Science, 150,* 367.

Galton, F. (1883) *Enquiries into Human Faculty,* London: Macmillan.

Goulding, A., Westerlund, J., Parker, A., & Wackermann, J. (2004). The first Digital Autoganzfeld study using a real-time judging procedure. *European Journal of Parapsychology*, *19*, 66–97.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, *29(2)*, 147–160.

He, M.X., Petoukhov, S.V., & Ricci, P.E. (2004). Genetic code, Hamming distance and stochastic matrices. *Bulletin of Mathematical Biology*, *66*, 1405–1421.

Honorton, C. (1978). Psi and internal attentional states: information retrieval in the ganzfeld. In B. Shapin & L. Coly (Eds.), *Psi and States of Awareness.* New York: Parapsychological Foundation.

Honorton, C. (1985a). How to evaluate and improve the replicability of parapsychological effects. In B. Shapin, & L. Coly (Eds.), *The Repeatability Problem in Parapsychology* (pp. 238–255). New York: Parapsychological Foundation.

Honorton, C. (1985b). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49*, 51–91.

Honorton, C., Berger, R., Varvoglis, M., Quant, M., Derr, P., Schechter, E. I., & Ferrari, D. C. (1990). Psi communication in the ganzfeld. *Journal of Parapsychology*, *54*, 99–139.

Honorton, C., Ferrari, D. C., & Bem, D. (1992). Extraversion and ESP performance: A meta-analysis and a new confirmation. In L. A. Henkel & G. R. Schmeidler (Eds.), *Research in Parapsychology 1990*, (pp. 35–38), Scarecrow Press.

Lantz, N., Luke, W., & May, E. (1994). Target and sender dependencies in anomalous cognition experiments. *The Journal of Parapsychology*, *58*, 285–302.

May, E., Spottiswoode, S. J. P., & James, C. L. (1994). Managing the target pool bandwidth: possible noise reduction for anomalous cognition experiments. *Journal of Parapsychology*, *58*, 303–313.

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

Parker, A., Grams, D., & Petterson, C. (1998). Some further variables relating to psi in the ganzfeld. *Journal of Parapsychology*, *62*, 319–337.

Parker, A., & Westerlund, J. (1998). Current research in giving the ganzfeld an old and a new twist. In *The Parapsychological Association 41th Annual Convention: Proceedings of Presented Papers*, (pp. 135–142). New York, NY: Parapsychological Association.

Parker, A. (2006). A ganzfeld study with identical twins. In *The Parapsychological Association 49th Annual Convention: Proceedings of Presented Papers*, (pp. 330–343). New York, NY: Parapsychological Association.

Playfair, G., L. (1999). Identical twins and telepathy. *Journal of the Society for Psychical Research*, 63, 86–98.

Pütz, P., Braeunig, M., & Wackermann, J. (2006). EEG correlates of multimodal ganz-

feld induced hallucinatory imagery. *International Journal of Psychophysiology*, *61*, 167–178.

Wackermann, J., Pütz, P., Büchi, S., Strauch, I., & Lehmann, D. (2002). Brain electrical activity and subjective experience during altered states of consciousness: ganzfeld and hypnagogic states. *International Journal of Psychophysiology*, *46*, 123–146.

Watt, C. (1988). Characteristics of successful free-response targets: Theoretical considerations. In *The Parapsychological Association 31st Annual Convention: Proceedings of Presented Papers*, (pp. 247–263). New York, NY: Parapsychological Association.

Weber, E. (1980). *Grundriss der biologischen Statistik*, 8th ed., Jena: Fischer.

Appendix

One of the two pairs producing three CTIs in their experimental session ('hat-trick') were female, 22 years old twins. The sisters showed high similarity in their appearance and habitus and looked like identical twins. However, the assumption has not been tested and we cannot say with certainty if they are mono-zygotic twins. About one year later, they contacted the experimenters and expressed their interest in participating in another ganzfeld experiment. A second experiment was carried out with the pair, using the same procedure and instructions as in the first one. However, we have to assume that the participants were at that time aware of the proper intent of the experiment. The second session was arranged with changed roles ($A \leftrightarrow B$) of the participants. Results of the second session are reported here for the sake of completeness, but they were not included into the data of the present study.

The similarity scores given by subject A in the second experiment yielded two correct target identifications. This result *per se* is not remarkable: the probability to obtain, by chance, at least two CTIs in three trials is $p = .156$. However, combining the two sessions and evaluating the probability to obtain at least five CTIs in six trials is $p = .0046$ (as given by binomial distribution $\mathbf{B}_6(.25)$), which is quite impressive.

Also remarkable are the subject A's ratings themselves (Table 6). The scores assigned to the correctly identified target clips were '99' and '100'. It is unlikely that these scores really respond to the experimental task, that is, to evaluate similarity between the ganzfeld-induced experience and the visual material (clip); they may rather reflect the subject's intention to indicate the target clip. In other words, the use of the extreme scores on the similarity scale corresponds to the shift from a 'covert' to the 'overt' experimental task, in which subject A attempted a correct target identification.

Table 6: Rating scores of twins in first and second session.

First session	Target	Decoy 1	Decoy 2	Decoy 3
Trial 1	89	19	28	11
Trial 2	83	15	68	75
Trial 3	82	77	56	29
Second session				
Trial 1	99	75	25	0
Trial 2	100	60	0	10
Trial 3	50	0	90	16

Of course, the single case reported here is merely suggestive of anomalous cognition and not a 'statistical proof'. Nevertheless, the idea that there may be 'special bonds' between twins is wide-spread; this not only as a popular belief but also as a topic of serious studies, pioneered by F. Galton more than a century ago (Galton, 1883, pp. 226–231). Also, these special ways of communication may be not restricted to anomalous cognition. For example, Duane and Behrendt (1965) described 'extra-sensory electroencephalographic induction' in two out of fifteen identical twins: oc-currence of EEG alpha rhythms in one subject reportedly 'induced' alpha rhythms in EEG of the other subject. In spite of an amount of literature on the topic of 'twin-telepathy' (see Playfair, 1999, for a review), the results are still inconclusive. Recently Parker (2006) reported preliminary results from a ganzfeld study with identical twins:

ten pairs of fifteen tested so far obtained a 'hit-rate' of 40% (in regard of the small sample size not significant). The question whether twins really are more likely to establish anomalous communication remains still open.

Psi as Compensation for Modality Impairment — A Replication Study Using Sighted and Blind Participants

Lance Storm and Mikele Barrett-Woodbridge

Anomalistic Psychology Research Unit, School of Psychology,
University of Adelaide, Adelaide, Australia

Abstract

A replication study of an earlier study by Storm and Thalbourne (2001) was conducted to test the hypothesis that blind people compensate for their impairment by developing superior psi ability compared to sighted people. Participants had to describe a concealed line drawing (target), and then rank four drawings (1 target + 3 decoys) from 'most likely' to be the target to 'least likely'. The concealed picture was removed from its envelope and assigned its corresponding rank. A significant psi effect was found for the whole sample, and for the sighted sub-sample, but not the vision-impaired sub-sample. An above-chance success-rate of 28% ($\pi = .54$, where $\pi_{MCE} = .50$) was found for the totally blind, which was superior (not significantly) to the rest of the sample (i.e., sighted + partially sighted participants) with their hit-rate of 26% ($\pi = .51$). In the present replication study, it was hypothesized that totally blind individuals have superior psi test performance to sighted individuals. However, the totally blind group and the sighted group both scored at the same below-chance hit-rate of 21% ($p = .365$; $\pi = .45$). There was thus no evidence that psi compensates for total blindness. When the dataset from the present study was combined with Storm and Thalbourne's (2001) dataset (total $N = 160$), the sighted group scored significantly above chance on the sum-of-ranks measure ($p = .040$). It was argued that if there is compensation for blindness, it might work in ways other than paranormal.

Correspondence details: Lance Storm, Anomalistic Psychology Research Unit, School of Psychology, University of Adelaide, Adelaide, SA 5005, Australia. Email: lance.storm@psychology.adelaide.edu.au.

Introduction

As far back as 1891, blindness, imagery and anomalous (ostensibly paranormal) performance have been topics of interest for parapsychologists (Alvarado, 1988). F. W. H. Myers (1891) believed that paranormal ability might manifest in the blind, and he suggested that the blind person:

> "...will exercise a sight, which he [*sic*] does not recognize as sight, which belongs in fact to that pre-natal undifferentiated continuum of perceptive faculty of which telepathic and clairvoyant phenomena show us the vestigial or obsolescent trace." [1]

> (F. W. H. Myers, 1891, p. 127)

It was not until the 1930s that Price and Pegram (1937) investigated psi performance in the vision-impaired, a group which included partially vision-impaired individuals. Using Zener cards (25 cards with five each of five Zener symbols: Star, circle, square, cross and wavy-lines), Price and Pegram administered runs of 25 calls to participants using three matching techniques: Open Matching (cards in the pack are face down and the participant sorts them into 1 of 5 identified key cards), Blind Matching (same as Open Matching, but the key cards are not identified), and Match Piling (the participant divides the 25 cards into 5 piles, and names the piles, e.g., "this is the circle pile, that is the star pile," etc.). Price and Pegram found that age and extent of impairment did not make significant differences to scoring trends. Even more surprising was the fact that scores actually improved when the cards in the pack were placed in opaque sealed envelopes.

Price and Pegram (1937) recognized the problem of testing a specific group without using a control group for comparative purposes (i.e.,

---

[1]One referee of the present article suggested that Myers did not mean that blind people would have their psi function sharpened due to loss of some other modality. We are not claiming he did. We merely take a lead from Myer's words, in order to run with the idea that vision-impaired people (especially those born blind) might, by default, maintain a close connection to a function that is related to psi in some way, and may be enhanced as a corollary of the fact that they are vision-impaired, whereas for sighted people the connection to the psi component of that function may be diminished as a consequence of undergoing a developmental process which, as a matter of convenience, we might call 'normal sight-training' that ultimately becomes the dominant sensory modality.

they did not test sighted people). They concluded that the "high proportion [of hits] cannot be considered significantly unusual until further research has been done with non-blind groups of comparable age and grade under similar social conditions" (p. 153). Consequently, in a follow-up study, Price (1938) tested vision-impaired and sighted participants so that performance comparisons could be made. The two groups "were selected for similarity on age and institutional status" (p. 286). There were 66 "blind" and 40 "seeing" participants. Open Matching and Blind Matching methods were primarily used, as in the Price and Pegram (1937) study. When both groups were combined, overall scoring was significantly above chance. The vision-impaired produced higher average scores than the sighted, but the differences were not significant. When cards were sealed in opaque envelopes, scoring was significantly higher for both groups compared to the 'open-card' method. Price concluded that "something meaningful" (i.e., "extra-sensory perception") took place in the tests, and she dismissed the rival hypotheses of sensory leakage that might explain the effects, as was shown by the fact that "subjects scored better in tests with enclosed cards than with the open cards" (p. 282).

Gonzales-Scarano (1982) looked at paranormal task performance in the sighted and the vision-impaired. She (after Paivio, 1971) assumed compensation in those who suffered from impairment in one of the normal modalities, and she theorized that unconscious visual and/or auditory images would be activated into consciousness by noncognitive factors, which could even be paranormal in nature. A visual memory test similar to that of Paivio and Okovita (1971) was used. High-visual/low-auditory word-pairs and high-auditory/low-visual word pairs were randomized and presented verbally to the participants. The paranormal component of the test was described as involving the identification of pre-designated specific word targets. The experimenter was 'blind' to the identity of the targets. Two hypotheses were proposed: (i) the sighted would recall more 'high-visual concrete nouns' than the congenitally totally blind, and (ii) the congenitally totally blind would recall more 'high-auditory concrete nouns' than the sighted. Both hypotheses failed to be confirmed. These results suggest that the totally blind participants were no more advantaged or disadvantaged than the sighted.

Like Price (1938), Barnard and Nelson (1983) also ran a card-matching task using 10 sighted and 10 "nonsighted" participants. They hypothesised that the nonsighted would perform significantly better

than the sighted, especially when the nonsighted were allowed to "touch" (i.e., "directly handle," p. 58) the cards, rather than not touch them. There were therefore four groups, but no significant main effect, or interaction effect, was found. However, a significant variance difference was found (as a so-called F max ratio of the variances) between the sighted and nonsighted groups. Also, in a *post-hoc t*-test comparing overall group scores to chance, significant "psi hitting" was found for the "nonsighted" (p. 59). The combined sample of 20 participants also produced a significant hit rate, but only in the 'touch' condition.[2]

Overall, the few available studies show that there is only a suggestion that performances of the vision-impaired may be superior to those of the sighted. Perhaps, more importantly, Gonzales-Scarano's (1982) negative findings suggest that the vision-impaired and the sighted should not be treated as incommensurable groups when given tasks that test their capacities to form images in their minds. There may be no good reason not to compare the vision-impaired and the sighted.

Storm and Thalbourne's (2001) experiment

Thalbourne's (2004) "diasomatic hypothesis" proposes that paranormal processes are seen as acting both "inside and outside the body" (Thalbourne, 2003, p. 31). This hypothesis emerges from his *Theory of Psychopraxia*, where psychopraxia is defined as:

> "A ...principle underlying all interactions between the self, or ego, and the realm consisting of mental and physical events, whereby *under certain conditions* ...the adoption of a pro attitude ...results in its fulfillment in reality. Paranormal phenomena may thus be seen as special instances of psychopraxia, being those manifestations of goal-achievement which are exosomatic rather than endosomatic, i.e., which are not mediated by the normal sensory-motor apparatus."

> (Thalbourne, 2003, p. 100)

---

[2]The other referee of this article suggested that psi hitting in the "touch" condition might be an artefact of tactile contact (i.e., "handling the stimuli"). Though possible, we would have to make the unlikely assumption that the card-symbols, printed in ink on cardboard cards, could be felt through the thick envelope. We would suggest that tactile contact is merely psi-conducive because it creates a feeling of psychological propinquity that helps overcome the psychological distance between the symbol and the participant, whereas the non-touch condition may have a distancing effect.

Psychopraxia is thus the self (*psyche*) bringing about goals (*praxia*, from the Greek *prattein*: "to accomplish") endosomatically in the mind-body complex, or exosomatically in the wider world.

Exosomatic psychopraxia (or the self enacting an effect external to the body, otherwise referred to as a paranormal effect) acts either in a compensatory way (substitution mode) or as a 'spill-over' effect. Exosomatic psychopraxia can therefore be seen as a compensation for some temporary or permanent ostensibly 'adverse condition' (e.g., vision-impairment). Storm and Thalbourne's (2001) initial experiment with the vision-impaired (described below) was the first of its kind in which vision-impaired participants were administered a free-response task rather than a forced-choice (card-guessing) task, as was the case in Price and Pegram's (1937) study.

Storm and Thalbourne asked sighted and vision-impaired participants to describe verbally a concealed randomly selected line drawing. Participants then ranked, from 1 to 4, four pictures that were removed from another envelope and placed in front of them (one of the pictures was a copy of the target picture and the other three were decoys). Totally blind participants had the four pictures described to them by the experimenter since they could not see them. Rank-scores were analysed using the sum of ordinal weighted ranks formula (Solfvin, Kelly, & Burdick, 1978, p. 99). Storm and Thalbourne found that the sighted significantly out-performed the vision-impaired, but their experiment was less than ideal given that totally blind participants were not readily available so that the sub-sample of vision-impaired had to include *partially* blind participants, many of whom were elderly with vision problems at the time of testing, but nevertheless did have good vision for most of their lives.

In a *post-hoc* analysis, Storm and Thalbourne conducted an analysis of rank-scores for both vision-impaired subgroups (i.e., totally blind and partially blind). They found that 28% of the totally blind subgroup ($n = 18$) scored a rank of 1 (i.e., correct on the first guess) whereas only 13% of the partially blind subgroup ($n = 24$) scored a rank of 1. This direct-hits rate for blind participants translated as an effect size of $\pi = .54$ (where π is a 'proportional index' based on the proportion of direct hits),[3] which was actually greater than the effect size of $\pi = .51$ for the

---

[3] $\pi = \frac{P(k-1)}{[1+P(k-2)]}$ where P is the raw proportion of hits, and k is the number of alternative choices available. Bem and Honorton (1994) point out the advantage this measure has in providing a "straightforward intuitive interpretation" (p. 8) of the effect size, because π is the "proportion correct, trans-

remaining 66 sighted and partially blind participants (direct-hits rate: $p = .26$). Though the differences were not significant, Storm and Thalbourne conjectured that the sighted and partially sighted participants might have no advantage over the totally blind participants.

The replication experiment described in the present study used totally blind participants only, rather than partially blind participants, in order to test Thalbourne's compensation hypothesis more effectively. The major aim was to determine whether totally blind participants could use psi in a compensatory way that would result in performance superior to that of sighted participants.

Method

Participants

Total number of participants was 76. Mean age was 55 ($SD = 18$ years). The experimental group consisted of 38 totally blind participants (mean age = 55, $SD = 18$ years) randomly selected from vision-impaired communities with the assistance of the various institutes that represent this group.[4] The control group of sighted participants were drawn from the general population, and they were matched with the experimental participants on age and sex ($n = 38$; mean age = 55, $SD = 18$ years).

Procedure

After the School of Psychology, University of Adelaide, granted ethics approval to conduct the experiment, the project leader (L. S.) initiated contact with the management of the participating institutions that represented the vision-impaired communities. The experimental component of this study was conducted by M.B.W. Prior to testing, for each of 76 trials, L. S. randomly (i) selected a target picture (a hand-drawn image randomly selected from a dictionary)[5] from four similarly derived pictures, (ii) photocopied the picture, (ii) wrapped it in aluminium foil,

---

formed to a two-choice standard situation" so that $P_{MCE} = .50$ (Rosenthal & Rubin, 1989, p. 333).

[4]Vision-impaired participants were acquired with the assistance of South Australian institutions including Townsend House, the Royal Society for the Blind, the Blind Welfare Association, Guide Dogs Association, and Radio Station 5RPH. Interstate participants were acquired with the assistance of four Victorian institutions: Blind Citizens Australia, Royal Victorian Institute for the Blind, the Deaf Blind Association, and Vision Australia Foundation.

[5]The gallery of 180 pictures used in this experiment is comprised of Thalbourne's (1981) hand-drawn originals. Randomization was achieved by using Pagano's (1986, pp. 479–480, Table J) random number tables. Each picture was randomly assigned to a four-picture set, so that there were 45 sets altogether. Each set was then assigned a number and set selection was achieved by using Pagano's (1986) Table J.

and (iv) concealed it in a target envelope. M. B. W. thus remained 'blind' to the target during each trial.

All vision-impaired participants were tested in their homes. Participants' details (age, gender, and level of blindness) were recorded. Vision-impaired participants fell into two categories: (i) blind from birth, and (ii) blind after birth.

All participants completed (1) Thalbourne's (1995) Australian Sheep-Goat Scale (ASGS), (2) the Extraversion (EX) sub-scale from the EPI (Eysenck & Eysenck, 1965); and (3) Rosenberg's (1965) Self-Esteem (S-E) Scale. The EX Scale and the S-E Scale were administered because Storm and Thalbourne (2001) argued that vision-impaired participants seemed more introverted and lower in self-esteem than sighted participants; introversion and low self-esteem being possible psi-inhibitive factors. Regarding the psi-extraversion relationship, see the meta-analysis by Honorton, Ferrari, and Bem (1998). Note also that extraversion and self-esteem tend to correlate significantly and moderately (see Robins, Tracy, & Trzesniewski, 2001). Items from both scales were read out by M. B. W. to blind participants.

Participants were then required to describe verbally the line drawing that was concealed in aluminium-foil inside the manila envelope. The experimenter M. B. W. took notes of the participant's mentations, and read them back, in order to prompt the participant's memory, thereby assisting them in the ranking process. M. B. W. did not offer personal interpretations as that might have misled participants.

Participants ranked the four pictures from 1 to 4 (1 being the 'most likely' picture concealed in the envelope, 4 being the 'least likely'). M. B. W. gave impartial assistance in the judging process by describing the pictures to the blind participants since they could not see them. All participants received a payment of AUD$20.00 for volunteering. L. S. conducted all statistical tests.

The following hypotheses are proposed (tests used are also given):[6]

1. Blind participants score higher than sighted participants using: (i) the direct-hits measure (i.e., percent correct, where MCE = 25% (ex-

---

[6]Note that the various measures of psi in Hypotheses 1 and 2 are conducted for comparative purposes. Initially, Storm and Thalbourne (2001) used the sum-of-ranks test, whereas the direct-hits test was post hoc. Both types of tests are again conducted in the present study merely to pursue some conjectures about how psi might manifest in blind and sighted participants. Whilst the likelihood of Type I error is increased, we let the results speak for themselves, and we leave it open-ended as to whether we found evidence for psi (see Discussion for further comments).

act binomial test, one-tailed, and t test, one-tailed), and (ii) the effect size measure π, where $\pi = \frac{P(k-1)}{[1+P(k-2)]}$ (note that π_{MCE} = .50).

2. The levels of scoring, as sum-of-ranks scores for (i) the whole sample, (ii) the blind group, and (iii) the sighted group, are lower (i.e., better) than chance (MCE = 2.50). (Test used: the sum of ordinal weighted ranks formula, one-tailed).[7]

3. There is a difference in performance between the blind and the sighted such that the mean sum-of-ranks score for the totally blind is lower (better) than the mean sum-of-ranks score for the sighted (Test used: Wilcoxon signed-ranks matched-pairs test, one-tailed).

4. Performance of sheep will be superior to goats using mean rank scores (based on a median-split division of ASGS scores). (Test used: Mann-Whitney U test, one-tailed).

5. There are negative relationships between rank-scores and (i) ASGS scores; (ii) extraversion, and (iii) self-esteem (Tests used: Spearman's rho test, one-tailed).

6. There is a positive relationship between extraversion and self-esteem (Test used: Pearson's r test).

Results

Descriptive data

Frequencies of 'hits' by ranks for each group and the whole sample are given in Table 1.

Table 1: Participants' Rank Scores: Totally Blind, Sighted, and Whole Sample

Group	Rank				Total
	1	2	3	4	
Totally Blind	8 (21.1%)	7 (18.4%)	12 (31.6%)	11 (28.9%)	38
Sighted	8 (21.1%)	12 (31.6%)	9 (23.7%)	9 (23.7%)	38
Whole Sample	*16 (21.1%)*	*19 (25.0%)*	*21 (27.6%)*	*20 (26.3%)*	*76*

---

[7]Level of scoring is determined from the sum-of-ranks score and the corresponding Z score. $Z = \frac{(M-U_M \pm 0.5)}{\sigma_M}$, "where M is the observed sum-of-ranks, $U_M = \frac{N(R+1)}{2}$, and $\sigma_M^2 = \frac{N(R-1)}{12}$. The 0.5 is the usual continuity correction and has sign opposite to that of $(M - U_M)$" (see Solfvin, Kelly, & Burdick, 1978, p. 99). Psi-hitting is indicated by a significant sum-of-ranks score that is lower (better) than MCE = 2.50. The Z score will be *negative* because U_M is greater than M.

The mean Australian Sheep-Goat Scale (ASGS) score for the whole sample was 19.71 (SD = 7.07). The difference between ASGS scores for blind (M = 19.21, SD = 6.44) and sighted (M = 20.21, SD = 7.71) was tested, but the result was not significant, $t_{(74)}$ = -0.61, p = .541 (two-tailed).

The mean EX score for the whole sample was 14.03 (SD = .28). The difference between EX scores for blind (M = 13.76, SD = 4.49) and sighted (M = 14.29, SD = 4.12) was not significant, $t_{(74)}$ = -0.53, p = .596 (two-tailed).

The mean S-E score for the whole sample was 19.16 (SD = 4.29). The difference between scores for blind (M = 19.32, SD = 5.56) and sighted (M = 19.00, SD = 5.82) was not significant, $t_{(74)}$ = 0.24, p = .810 (two-tailed).

Planned analyses

Hypothesis 1: It was hypothesised that blind participants would score higher than sighted participants using the direct-hits measure as a proportion-correct, and effect size π measure. The totally blind group (with 8 hits) and the sighted group (also with 8 hits) both scored at the below-chance hit-rate of P = 21%, exact Binomial p = .769 (π = .45, z = -0.51, p = .829, right-tailed). There was no evidence that psi compensates for total blindness. Note that these two identical hit-rates indicate that the whole sample also scored at P = 21%, exact p = .822 (π = .45, z = -0.72, p = .764, right-tailed).

Hypothesis 2: It was hypothesised that the levels of scoring, as sum-of-ranks scores for (i) the whole sample, (ii) the blind group, and (iii) the sighted group, would be lower (better) than chance. The following results were obtained:

1. Whole sample: z = 0.67, p = .749 (left-tailed). The mean rank score was 2.59 (SD = 1.10), which was greater (i.e., worse) than chance, where MCE = 2.50.

2. Totally blind: z = 0.94, p = .826 (left-tailed). The mean rank score was 2.68 (SD = 1.12), which was greater (i.e., worse) than chance.

3. Sighted: z = 0.00, p = .500 (left-tailed). The mean rank score was exactly at chance 2.50 (SD = 1.08).

None of the results were in the expected directions, and none were significant.

Hypothesis 3: It was hypothesised that there would be a difference in performance between the blind and the sighted such that the mean rank score for the totally blind would be lower (better) than the mean rank score for the sighted. The mean rank score for the totally blind (M = 2.68) was not superior to that of sighted participants (M = 2.50). The hypothesis was not supported. Since the hypothesis is directional (one-tailed), the difference was not tested.

Hypothesis 4: It was hypothesised that the performance of sheep would be superior to goats using mean rank scores (based on a median-split division of ASGS scores). The median ASGS score was 20. Those above 20 were taken as sheep (n = 36), those below or equal to 20 were taken as goats (n = 40). Performance by sheep (M = 2.72 SD = 1.16) was not superior to that of goats (M = 2.48, SD = 1.04). The hypothesis was not supported.

Hypothesis 5: It was hypothesised that there would be negative relationships between rank-scores and (i) ASGS scores, (ii) extraversion, and (iii) self-esteem. All three relationships were positive. Relationships were also positive for the same three correlations when data for only the totally blind were used. The hypothesis was not supported.

Hypothesis 6: It was hypothesised that there would be a positive relationship between extraversion and self-esteem. The relationship was positive, and it was significant, $r_{(74)}$ = 0.23, p = .023 (one-tailed). The hypothesis was supported.

Post-hoc analyses

Performance Comparisons: Given the inferior non-significant overall performance of the sample compared to the overall significant performance of Storm and Thalbourne's (2001) sample, the mean rank scores of the two datasets were tested for homogeneity. If there is a significant difference, then the two samples were not drawn from the same population. The Mann-Whitney U test is used to test the difference. The test result was not significant, U = 2683.5, Z = -1.80, p = .072 (two-tailed). The

merging of the two samples is therefore justified, given that the result suggests the samples are not heterogeneous.

Scoring on the Direct-Hits Measure: Hypothesis 1 was re-tested using the dataset from the present study ($N = 76$), combined with Storm and Thalbourne's (2001) dataset ($N = 84$: Total $N = 160$). Results are given in Table 2. The totally blind (23.2%) did not score higher than the sighted (27.5%). Again, no support was found for Hypothesis 1. Storm and Thalbourne (2001) initially found that the totally blind scored better than the rest of the sample. The totally blind ($P = 23.2\%$) did not score better than the rest of the sample (i.e., sighted/partially-sighted, $P = 24.0\%$).

Table 2: Direct-Hit Rates: Various Groups and Whole Sample

Group	N	Hits	P (%)	p	π	z	p[a]
Totally Blind	56	13	23.2	.671	.47	-0.38	.648
Sighted	80	22	27.5	.343	.53	0.48	.316
Partially-Sighted	24	3	12.5	.960	.30	-0.92	.821
Sighted & Partially-Sighted datasets combined	104	25	24.0	.626	.49	-0.17	.433
Whole Sample	*160*	*38*	*23.8*	*.672*	*.48*	*-0.43*	*.666*

[a]p values for the z statistics are right-tailed.

Scoring on the Sum-of-Ranks Measure: The sum-of-ranks formula was applied to the combined datasets of Storm and Thalbourne (2001; $N = 84$) and the present study ($N = 76$) to re-test Hypotheses 2 and 3. Results are given in Table 3.

Table 3: Mean Rank Scores: Various Groups and Whole Sample

Group	Mean Rank	SD	z	p[a]
Totally Blind ($n = 56$)	2.57	1.11	0.42	.663
Sighted ($n = 80$)	2.27	1.04	-1.75	.040
Partially-Sighted ($n = 24$)	2.63	1.01	0.46	.677
Sighted & Partially-Sighted datasets combined ($n = 104$)	2.36	1.04	-1.27	.102
Whole Sample ($N = 160$)	*2.43*	*1.07*	*-0.74*	*.230*

[a]p values are left-tailed.

Only the sighted group produced a mean rank score that was significantly lower (better) than chance, which strengthens Storm and

Thalbourne's (2001) initial finding of a significant mean rank score for sighted individuals. Hypothesis 2 was therefore partially supported.

In the case of Hypothesis 3, the mean rank score of the totally blind ($M = 2.57$) was not found to be superior to that of the sighted ($M = 2.28$). The hypothesis was not supported. Since the hypothesis is directional (one-tailed), the difference was not tested.

Discussion

Planned analyses

In the present study, the blind group's performance was not superior to that of the sighted group on the direct-hits measure. In fact, both groups got the same below-chance score of 21%, where MCE = 25%. However, using the sum-of-ranks formula, the sighted group scored a little better than the totally blind, but not significantly better, and once again, scoring was no better than chance. These results are of some concern, not because the differences were not significant, but because of the disconcerting fact that both groups scored comparatively worse than their corresponding cohorts in Storm and Thalbourne's (2001) initial study on both measures (i.e., direct hits and sum-of-ranks, see Storm & Thalbourne, 2001, pp. 151-153).

Three measures — Thalbourne's Sheep-Goat Scale, Eysenck's Extraversion scale, and Rosenberg's Self-Esteem scale — were administered to the sample to determine if rank scores could be predicted from scores on those measures. For the four tests conducted, there were no significant correlations between rank scores and any of the three measures, although (perhaps not surprisingly) extraversion correlated positively and significantly with self-esteem.

Post-hoc analyses

The situation did not improve much when using the direct-hits measure on the combined dataset (i.e., data from the present study ($N = 76$) plus the data from Storm and Thalbourne's study ($N = 84$) providing a total $N = 160$). For the combined sample ($N = 160$), direct hitting was not superior for the totally blind compared to the sighted. Direct hitting was also not superior for the totally blind group compared to the remainder of the sample (i.e., sighted/partially-sighted). The results do not lend any support to the hypothesis that total blindness is compensated by *enhanced* psi performance (i.e., psi hitting).

However, using the sum-of-ranks formula, an overall significant effect was found for the sighted group only. This result is not an entirely independent replication of the effect initially found by Storm and Thalbourne (2001) because the combined score (M = 2.27) uses high-scoring old data (i.e., Storm & Thalbourne's, 2001, data) to bolster up the exactly-at-chance score (M = 2.50) of the new sighted group. On the other hand, the result may more accurately reflect its corresponding population parameter for sighted individuals. From this result, it appears that sighted participants had an advantage over vision-impaired participants, who again scored worse than chance.

These results do not lend any support to the hypothesis that total blindness is compensated by *enhanced* psi performance (i.e., psi hitting). However, it is also true that any statement one might like to make about psi (as measured on the free-response task) for sighted individuals will depend on how conservative one thinks the psi measure should be. The direct-hits measure is said to be the most conservative (cf. Honorton, 1985), but Solfvin et al. (1978) argue that all ranks should be considered in order that partial credit be given to the other ranks.

The problem of ideal targets

In considering the alleged compensation effect in the vision-impaired, it may be the case that compensation comes in other forms other than enhanced psi performance. This alternative hypothesis can be tested in ways too numerous to mention here, but we imagine testing would focus primarily on the still-functioning modalities such as hearing and touch (see Sacks, 2003). Oliver Sacks discusses a hearing compensation effect in John Hull, who became totally blind at the age of 13. Of Hull's improved hearing, Sacks writes:

> "With his new intensity of auditory experience (or attention), along with the sharpening of his other senses, Hull comes to feel a sense of intimacy with nature, an intensity of being-in-the-world, beyond anything he knew when he was sighted. Blindness now becomes for him 'a dark, paradoxical gift.' This is not just 'compensation,' he emphasizes, but a whole new order, a new mode of human being."
>
> (Sacks, 2003)

However, our issue is with ostensibly paranormal (not normal) experience. Furthermore, to shift our focus somewhat, and in spite of our

earlier statements to the contrary (cf. Gonzales-Scarano, 1982), we now find ourselves in a quandary over target suitability. It is possible that hand-drawn pictures may, in some way, facilitate the psi process more effectively in sighted individuals simply because they can visually scan and contemplate the four-choice array of drawings in front of them, thus triggering unconscious psychic processes that may subsequently result in a psi effect. The post hoc sum-of-ranks results suggest that this is the case for sighted individuals. Clearly, the totally blind may well be disadvantaged if vision is conducive to the psi process. Thus, we ask the question, What constitutes an ideal target given a participant's disability? It is still true that totally blind people may find it difficult to relate to the idea of a picture-target drawn on a piece of paper, in the sense that such targets are outside their normal day-to-day experiences, and this is especially true for individuals who are totally blind from birth. In the case of John Hull, for example, images and the very idea of seeing (the appearance and concept of an object once known to him) was lost forever over time: it was as if Hull had been blind his whole life (Sacks, 2003).

Broughton (1976) has pointed out that there may be target preferences amongst participants, dependent upon their natural inclinations and dispositions. For example, they might prefer three-dimensional or textured targets made of wood or other material. Broughton used five three-dimensional targets comparable to the five Zener cards in a design where participants used either their left or right hands to select the shape that 'felt correct'. Other possible tests for blind participants might include musical notes or sound segments (cf. Willin, 1996), taste recognition, and odour discrimination (cf. Stahl, 2004). It seems to us, however, that targets of this nature, and contact with those targets, would be psi-conducive during the ranking stage more than anything else if we are to assume that the psi process is democratic. After all, surely the blind person should not be allowed to make tactile contact with the target set during the mentation stage; the sighted person is never given a similar advantage since the four pictures are concealed in an envelope until the ranking stage.

We must also consider whether giving totally blind people a sensory advantage does not in itself give sighted people a disadvantage. Just as there is differential functioning between the hemispheres of the brain such that a person with left-brain dominance may prefer language-based targets (e.g., words, phrases, sentences) or sequential

targets (e.g., ordered pattern changes), a person with right-brain dominance may prefer visual or pictorial targets. Clearly, our aim would be to devise targets that are neutral across the comparison groups.

As it stands, our findings may still be important if we have discovered something new about how the psi function might work. Not only should target suitability be a consideration when conducting experiments of this kind, but also we must recognise that all participants are not born equal, such that the psi function (insofar as it may be 'gestalt' in nature) may not necessarily be a democratic process *if its function is partially dependent on one or more normal sensory modalities*.[8]

Conclusion

For the new data, it was shown that the totally blind and the sighted are equally matched performance-wise in the picture-identifying task: scoring was at chance for both groups (i.e., no psi effects were found). Thus, there were no replications of Storm and Thalbourne's earlier findings. However, for the combined dataset, while only chance scoring was indicated in the vision-impaired groups, psi hitting *was* demonstrated in the sighted group. It appears that vision-impairment is not conducive to psi hitting. We argue that it is reasonable to assume that if there is compensation for vision-impairment, it might work in ways other than paranormal, but until tests are conducted using alternative targets that are preferable to both groups — sighted and vision-impaired — we are still not in a position to make inferences about psi compensation for modality impairment: specifically, total blindness.

Acknowledgements

---

[8]We acknowledge that the second referee of the present article made similar criticisms of our research. However, we are not just pointing out putative deficits in our study for the mere sake of it. We are trying to bring attention to a possible differential psi effect that may be governed by at least two factors: (i) the nature of the target, and (ii) how the target is mediated during the ranking stage.

References

Alvarado, C. S. (1988). ESP and imagery in the blind: Some early speculations. *Parapsychology Review, 19*, 5, 4-5.

Barnard, B. N., & Nelson, S. M. (1983). The nonsighted versus the sighted in ESP testing using card-matching as the decisive factor. Abstract in *Journal of Parapsychology, 47*, 58-59.

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

Broughton, R. S. (1976). Possible brain laterality effects on ESP performance. *Journal of the Society for Psychical Research, 48*, 384-399.

Eysenck, H. J., & Eysenck, S. B. G. (1965). *The Eysenck Personality Inventory*. San Diego: Educational and Industrial Testing Service.

Gonzales-Scarano, F (1982). ESP and imagery in the sighted and the blind. Unpublished Master's thesis, John F. Kennedy University, Orinda, CA.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49*, 51-91.

Honorton, C., Ferrari, D. C., & Bem, D. J. (1998). Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology, 62*, 255-276.

Myers, F. W. H. (1891). The principles of psychology. *Proceedings of the Society for Psychical Research, 7*, 111-133.

Pagano, R. R. (1986). *Understanding statistics in the behavioural sciences* (2nd ed.). New York: West Publishing.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A., & Okovita, H. W. (1971). Word imagery modalities and associative learning in blind and sighted subjects. *Journal of Verbal Learning and Learning Behavior, 10*, 506-510.

Price, M. M. (1938). A comparison of blind and seeing subjects in ESP tests. *Journal of Parapsychology, 2*, 273-286.

Price, M. M., & Pegram, M. H. (1937). Extra-sensory perception among the blind. *Journal of Parapsychology, 1*, 143-155.

Robins, R. W., Tracy, J. L., & Trzesniewski, K. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35*, 463–482.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106*, 332-337.

Sacks, O. (2003). *The mind's eye: What the blind see.* Retrieved February 3, 2006 from http://geoffandwen.com/Blind/newsarticle.asp?u_id=1218.

Solfvin, G. F., Kelly, E. F., & Burdick, D. S. (1978). Some new methods of analysis for preferential-ranking data. *Journal of the American Society for Psychical Research*, 72, 93-114.

Stahl, S. (2004). Permutation-based methods for examining confusion data in ESP experiments. *Journal of Parapsychology, 68*, 381-403.

Storm, L., & Thalbourne, M. A. (2001). Paranormal effects using sighted and vision-impaired participants in a quasi-ganzfeld task. *Australian Journal of Parapsychology, 1*, 133-170.

Thalbourne, M. A. (1981). Some experiments on the paranormal cognition of drawings, with special reference to personality and attitudinal variables. Unpublished Ph.D. thesis, University of Edinburgh, Scotland.

Thalbourne, M. A. (1995). Further studies of the measurement and correlates of belief in the paranormal. *Journal of the American Society for Psychical Research, 89*, 234-247.

Thalbourne, M. A. (2003). *A glossary of terms used in parapsychology.* Charlottesville, VA: Puente.

Thalbourne, M. A. (2004). *The common thread between ESP and PK.* New York: The Parapsychology Foundation.

Willin, M. J. (1996). A ganzfeld experiment using musical targets. *Journal of the Society for Psychical Research, 61*, 1-17.

Book Review

John Barry

Department of Psychology,
University of Westminster, London, U.K.

A Review of "Remote Viewing: The Science and Theory of Nonphysical Perception", by Courtney Brown (2005)

Have you read Bentov's (1977) *"Stalking the Wild Pendulum: On the Mechanics of Consciousness?"* It is the first of his only two books, and is full of the most imaginative speculation on life, the universe, and everything, couched in terms of physics and spirituality, and must be considered a classic of its genre. Bentov died in a plane crash just two years after its publication, a tragedy that perhaps causes readers to make allowances for flaws in his work, a rose-tinted indulgence of the sort lent to James Dean, Jim Morrison and the like. For Bentov these privileges include never having to provide actual hard evidence for any of his fantastic musings, never having to disappoint fans, embarrass employers etc. No such privileges are availed to Courtney Brown, alive and well and author of *"Remote Viewing: The Science and Theory of Nonphysical Perception."* This highly thought-provoking account of his recent research and theorizing on the phenomenon of remote viewing will, no doubt, carry with it expectations of testable theories, replicable studies, etc.

Brown is the Director of the *The Farsight Institute* in Atlanta, Georgia. This is an organisation dedicated to researching, teaching, and promoting public awareness of remote viewing. Brown has a background in research in the social sciences and has published respected papers on non-linear modelling. His involvement with remote viewing has been more controversial, and it is a relief to find that his latest book

Correspondence details: John Barry, Department of Psychology, University of Westminster, Headquarters Building, 309 Regent Street, London, W1B 2UW, United Kingdom. Email: j.barry@westminster.ac.uk.

is mercifully circumspect in references to, for example, viewings of non-physical beings. *Remote viewing* itself is, in a nutshell, clairvoyance: e.g., being able to describe a monument in one city whilst sitting in a room in another, having been given only the coordinates of the monument. Incredibly, remote viewing defies time: e.g., being able identify the monument without being given coordinates and before the target monument has even been chosen. Explaining this confusing mix of clairvoyance and precognition is a major theme in this book.

Brown spends about 150 pages describing his research. Perhaps because of my background in teaching research methods in psychology, this was the part of the book I was most looking forward to. However, Brown states that this book is concerned with "sociophysics" — a quantum theory explanation of human perception — rather than psychology, which in part explains why much of the statistical analysis left me scratching my head. I bow to Brown's greater knowledge of all things statistical and mathematical, and found that my knowledge of statistical modelling was no aid in understanding his descriptions of mathematical modelling or the *Russell Procedure* he employs (neither of which appear in any of my statistics books). However there were parts of the book that I felt I should have easily grasped but that were poorly explained. For example, in the appendix for Chapter 6 the description of his use of chi-square — although detailed — could have been clearer. On the other hand his computerised scoring method seems a useful development on previous systems (e.g., at the *Princeton Engineering Anomalies Research* laboratory). Also the use of computerised analysis is something to be applauded, as it would minimise any psychic contamination of results by a human analyst. Again precognition is an issue, as the thoughts of the person who initially decides whether what the remote viewer described matches the target(s) may inadvertently influence the choice of target that the remote viewer made in the first place. Overcoming this interesting experimenter effect is another major theme of this book, and leads to intriguing conclusions in the latter chapters regarding the action of the mind in what Brown calls "subspace", which can be thought of as the place where the mind travels to in remote viewing. These conclusions include *Brown's Rule*, which suggests an inverse relationship between remote viewing success and the probability of remote viewing success — e.g., remote viewing is more successful when the viewer is choosing one target from a pool of 500 rather than one target from a pool of two.

In the later chapters of this book Brown theorises on the nature of remote viewing in terms of quantum theory. His description of entanglement and the famous 'two-slit' experiment are very clear. However, when these are tied into his explanation of remote viewing things become less clear, but this is perhaps due to my limited knowledge of things quantum. It wasn't until the final chapter that clarity is restored, with a brief but engaging speculation on the relationship between remote viewing and other phenomena including near-death experiences, the soul, and free will.

In her review of research into remote viewing (and Ganzfeld), Utts (1996) suggested that given the weight of evidence in favour of the existence of remote viewing that future research should focus on exploring the basic mechanisms involved. Brown is to be applauded for his efforts in this direction, and this book will help clarify useful future directions for research in this area.

Publication Details*: Brown, C. (2005). Remote Viewing: The Science and Theory of Nonphysical Perception. Farsight Press. 344 pp. ISBN-10: 0976676214, ISBN-13: 978-0976676218. Publication Price: £13.50 approx. (paper).*

References

Bentov, I. (1977). *Stalking the wild pendulum.* London: Wildwood House.

Utts, J. M. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration, 10,* 3-30.