THE RELATIONSHIP BETWEEN LABILITY AND PERFORMANCE AT INTENTIONAL AND NONINTENTIONAL VERSIONS OF AN IMPLICIT PMIR-TYPE PSI TASK

Hitchman, G. A. M. , Roe, C. A. and Sherwood, S. J.

ISSN 0022-3387

THE RELATIONSHIP BETWEEN LABILITY AND PERFORMANCE AT INTENTIONAL AND NONINTENTIONAL VERSIONS OF AN IMPLICIT PMIR-TYPE PSI TASK

By Glenn A. M. Hitchman, Chris A. Roe, and Simon J. Sherwood

ABSTRACT: A number of theories of psi such as Stanford's psi-mediated instrumental response (PMIR) model suggest psi can function without a person's awareness, and that their intent to exhibit psi may be counterproductive. However, few parapsychological studies have directly compared participants' performance at intentional and nonintentional versions of equivalent tasks. This study sought to address this issue whilst exploring the role of lability, suggested by Stanford to be predictive of a person's propensity to respond to extrasensory stimuli. 50 participants took part in both intentional and nonintentional versions of a 10-trial, binary, forced-choice precognition task. A contingent outcome task system involving positive pictures as reward for hit trials and negative pictures as punishment for miss trials was administered on a trial-by-trial basis. Participants scored marginally fewer hits than the mean chance expectation in both versions of the task, with no tangible difference in their performance between tasks. Furthermore, no relationship was found between the number of precognitive hits they achieved and their scores on a composite psychometric measure of lability, nor its constituent elements. However, participants' expectations that their luck could aid their performance, as well as their emotional reactivity, were significantly positively related to their tacit psi scores.

Throughout the history of parapsychological research, experimenters have used both intentional and nonintentional methods in their attempts to capture psi phenomena. Whilst certain intentional means such as the ganzfeld approach (Bem & Honorton, 1994) have, on occasion, yielded successful outcomes, some theories of psi such as Stanford's (1974, 1977, 1982, 1990) psi-mediated instrumental response (PMIR) model suggest that the wilful intent to produce extrasensory effects may actually hinder the underlying process. A series of recent studies by Luke and colleagues utilised a nonintentional precognition protocol that was consistent with Stanford's conceptualisation of psi as primarily an unconscious process that functions in the service of an organism's needs by activating pre-existing behaviours in response to threats or opportunities in the environment (Luke, Delanoy & Sherwood, 2008; Luke & Morin, 2009; Luke, Roe & Davison, 2008). The method, which was employed relatively consistently throughout the four studies, involved a picture preference task in which participants were asked to select a preferred image from a set of four fractal patterns. Participants were unaware that this was actually a covert test of precognition, as immediately after they indicated their favourite picture the computer would randomly pick one of the four fractal images as a target. Stanford's notion of psi as a goal-oriented process was reflected by a contingent outcome design. At the end of a 10-trial session, participants who had scored more hits than the mean chance expectation (MCE) of 2.50 were rewarded by being able to rate positive images (either erotic pictures aligned to their sexual preference or humorous cartoons), whereas those who had achieved fewer hits than MCE were punished by having to take part in a boring number vigilance task. In each of these studies, participants' mean hit rates exceeded MCE, and three of the four studies yielded independently significant evidence of a nonintentional precognition effect. The combined mean hit rate across all four studies was 2.92 (SD = 1.46), significantly greater than the MCE of 2.50, $t(197) = 4.04$, $p = .000078$, two-tailed), with an effect size (ES) of $r = .28$. Please note that throughout this paper, effect sizes for t tests are calculated according the following formula: $ESr = \frac{\sqrt{t^2}}{t^2 + df}$.

In addition to describing the functionality of psi, Stanford's PMIR model also makes claims regarding various situational and individual difference factors which may either facilitate or constrain psi from occurring. The studies by Luke and colleagues utilised associated questionnaire methods as rudimentary attempts to assess the impact of two such factors—latent inhibition and lability. Latent inhibition can be understood as the general tendency of an organism to ignore or filter out information from further cognitive processing that it has learned is irrelevant to its ongoing situational concerns (Lubow, 1989). It was assessed indirectly via Goldberg's (1999) measure of openness to experience. Meanwhile, the lability construct (defined more precisely below) is often used as an antonym for, or on a continuum with, stability, and the linear and nonlinear subscales of the Creative Cognition Inventory (Holt, 2002) were used as a convenient proxy measure. Following the assumption that psi may function via extrasensory information transfer, it was predicted that those with higher levels of latent inhibition would be more prone to filtering out extrasensory data at an early stage and hence show diminished performance on the precognition task relative to those with lower levels of latent inhibition (see Holt, Simmonds-Moore, & Moore, 2007). Similarly, it was predicted that more labile individuals would have a greater propensity to respond to subtle extrasensory biases within their cognitive systems and would therefore perform better at the precognition task than more stable individuals.

In the Luke et al. (2008) study, a significant positive relationship was observed between openness to experience and precognition scores, $r = .46$, $p = .01$, two-tailed. However, this correlation was not replicated in the subsequent study by Luke and Morin (2009; $r = -.08$, $p = .64$, two-tailed). Meanwhile, when considering the relationship between tacit psi scores and the lability construct, Luke and Morin (2009) failed to find significant correlations between precognitive performance and either of the subscales of the Creative Cognition Inventory (linear subscale: $r = .25$, $p = .17$; nonlinear subscale: $r = .20$, $p = .27$). Despite the inconclusive results regarding the covariates of tacit precognition, Hitchman, Roe and Sherwood (2012) were sufficiently encouraged by the overall success of the Luke and colleagues' protocol to attempt to replicate and extend the paradigm. Whilst the core facets of the method were preserved, the design of their study was refined in several ways, including the experimental software program being completely rewritten in an updated programming language to overcome fears that previous results may have been due to an artifact within the code, and the number of trials being increased from 10 to 15 to enhance the statistical power of the study. They did, however, retain all of the questionnaire measures that had been used throughout Luke and colleagues' four studies.

Participants in the Hitchman et al. (2012) replication scored more hits on the nonintentional precognition task than the MCE of 3.75 (mean hit rate = 4.02) but this difference was not significant, $t(49)= 1.14$, $p = .13$, one-tailed. Meanwhile, this study provided some indirect support for the assumed role of latent inhibition via a medium-sized positive correlation between the number of precognitive hits participants achieved and their scores on Goldberg's (1999) openness to experience scale, $r = .29$, $p = .02$, one-tailed. However, little evidence was found to indicate that lability could influence precognitive performance, as participants' nonintentional precognition hit rates were unrelated to their scores on the linear and nonlinear subscales of Holt's (2002) Creative Cognition Inventory: linear subscale; $r = .14$, $p = .16$; nonlinear subscale; $r = .03$, $p = .41$; both one-tailed. However, in both cases it was argued that the measures used to assess these constructs were too indirect to allow any firm conclusions in relation to their effects. In particular, creativity constitutes a very limited proportion of lability, which is a much broader concept reflecting a wider range of facets relating to the readiness for change in an organism. It may be that the particular elements of lability that have been assessed as covariates of psi performance to date are not the most influential, whereas other components of the construct that have not yet been considered may play a more pertinent role in the psi process. It was therefore considered worthwhile in the present study to explore the relationship between performance at precognition tasks and a more comprehensive measure of lability, with particular attention to how it relates to the PMIR model.

Lability is a construct that was first popularised within parapsychology by Braud (e.g., Braud, 1980, 1981, 2002; Braud & Schlitz, 1983; Braud, Shafer & Mulgrew, 1983). According to Braud (1980, p. 1), lability represents "the ease with which a system can change from one state to another, the amount of 'free variability' in the system". Lability can thus be characterised as the a priori probability (ready capacity) that a system will change its state in a given situation. In this regard, Braud suggested that the brain-mind may, at times, be constrained by specific structural patterns. On these occasions, the brain-mind is said to be in an inert state in which it is resistant to change. Once such structures have been released, he proposes that the brain-mind may be at greater liberty to adapt or reorganise itself in relation to psi-relevant information or events (Braud, 2002). He therefore proposed that psi phenomena are more likely to manifest when individuals are characterised by a more labile (and hence less inert) state.

To illustrate this, Braud offered the concrete example of when a specific pattern of neuronal activity is required in the facilitation of a particular memory. If the neurons required for that pattern of neuronal activity are temporarily engaged in other structures or activities, the specific memory will be provisionally inhibited. Once those neuronal patterns become unstructured or deconstrained, the neurons required for the previously inhibited memory are made available, allowing for the memory to be triggered. The relevance of this particular example to the PMIR model is clear when we consider that one of the mechanisms through which Stanford (1990) claimed psi-mediated responses could be accomplished is via the triggering of pre-existing response mechanisms, which could include particular memory traces.

Despite the conceptual promise of lability, finding an appropriate measure of this construct is less straightforward. Braud et al. (1983) tended to focus on the notions of cognitive and perceptual lability. They measured cognitive lability via an assessment of the fluency of word associations, whereas perceptual lability was measured by assessing how frequently participants' perceptions of the Necker cube alternated between the two potential representations. More recently, Roe and Holt (2006; Holt & Roe, 2006) devised a broader measure of lability that combined various established psychometric measures believed to be indicative of an individual's lability. This composite measure was designed to include a variety of emotional, cognitive, physiological, neurological, and behavioural elements. Specifically, the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) was included, particularly as Openness to Experience, a constituent element, is thought to reflect an individual's willingness to engage with novel ideas and values, whereas Neuroticism is said to be indicative of emotional instability and a poor ability to control impulses, with both factors being particularly pertinent to the notion of lability.

The composite scale also included a measure of temporal lobe lability (Personal Philosophy Inventory; Persinger & Makarec, 1987). According to Persinger, those with greater levels of temporal lobe lability tend to be more impulsive, suggestible, emotionally sensitive, imaginative, and have a greater interest in philosophical ideas. He also suggests they may be more prone to psi experiences (Persinger, 1989). To account for lability of mood, two items on Mood Lability designed to screen for bipolar disorder (Akiskal et al., 1995) were included. This disorder is characterised by grand fluctuations of mood, indicating a highly labile state. Finally, following Braud's (1981) assertion that lability is related to novelty generation, two measures of creativity were included (Creative Cognition Inventory; Holt, 2007; Emotional Creativity Inventory; Averill, 1999).

These measures have subsequently been adjusted into a refined lability scale (Drennan, Roe, & Broughton, 2011) consisting of 71 items. This refined measure has an adequate level of internal consistency ($\alpha = .86$) with factor analysis revealing five main elements: Intuitive Cognition (26 items, $\alpha = .92$), Conceptual Cognition (18 items, $\alpha = .42$), Ego-Orientated Cognition (12 items, $\alpha = .78$), Emotional Interpretation (10 items, $\alpha = .71$) and Analytical Cognition (5 items, $\alpha = .79$). Whilst this refined and validated measure was not available at the time of the present study, given that its items were selected from the aforementioned composite elements, the composite scale may be considered a reasonable assessment measure, and it also enables further analysis of its constituent elements. Interestingly, these researchers found a medium-sized negative correlation between the composite lability scale and spontaneous psychokinetic experiences. This reflects the trend that the majority of research in relation to lability to date has focused on its relationship with psychokinesis, with little research having been conducted in relation to extrasensory perception. This study consequently provided an ideal opportunity to evaluate the role of lability in a PMIR-type precognition task.

In addition to exploring the covariates of psi predicted by Stanford in his PMIR model, this experiment also presented an opportunity to test one of the key predictions of the model by comparing intentional and nonintentional versions of the psi task. According to the PMIR model, psi can function without the conscious intention or awareness of the individual and any consciously generated thinking or cognitive constraints in relation to the need for psi in a life event or experimental situation may significantly diminish the possibility for psi to be manifested. We could therefore expect both intentional and nonintentional psi to be possible in principle, although the potential for intentional psi may be restricted by the cognitive interference associated with participants' awareness of the need to fulfil a psi task.

Few studies in the parapsychological literature have directly compared performance across intentional and nonintentional versions of equivalent or similar tasks. Rao and Davis (1978) performed an experiment designed to assess experimenter effects across nonintentional and intentional psi tasks with a limited sample of 11 female participants. The intentional psi task consisted of a word-based ESP test in which participants were asked to explicitly guess a series of English and Telugu (an unfamiliar language to the participants) target words that were concealed from their conventional sensory faculties. The nonintentional psi task required participants to rank 40 items from a mood adjective check list on a 4-point scale which were later compared against a list of randomly generated target numbers ranging from 1–4 for each adjective. The results of the study indicated a differential language effect in the intentional psi task, with participants scoring significantly higher on English words than Telugu words, but only for one of the experimenters. For the nonintentional psi task, it was found that participants scored significantly higher when they gave different mood ranks in the second of two experimental sessions compared with those who gave the same ranks. Furthermore, the number of mood items checked differently across the two nonintentional psi task sessions was found to correlate significantly with the differential between scores across the two languages in the intentional psi task. These findings would seem to indicate a relationship between participants' performance at intentional and nonintentional psi tasks, with participants who showed a greater tendency towards the differential language effect in the intentional psi task also performing better at the nonintentional psi task. It is interesting to note that the results of the nonintentional psi task are also indicative of a lability effect, with participants who changed their ranks from one session to the next showing heightened performance relative to those who were more rigid in their responses. Overall, however, a direct comparison of the relative strength of intentional and nonintentional psi in this study is limited by the different nature of the two tasks. The present study therefore included both intentional and nonintentional psi tasks of the same type in order to test the predictions of the PMIR model in relation to consciously generated thinking and the cognitive constraints associated with intentionality in a much more direct way.

Regarding the methodological considerations of the present study, it is noteworthy that in the studies involving the Luke and colleagues paradigm described above, participants took part in the contingent positive or negative outcome task (a reward system similar to the feedback mechanisms employed in studies of intentional psi) only at the end of completing a run of 15 trials. However, meta-analyses of forced-choice precognition studies indicate that there is a significant positive relationship between the degree and immediacy of feedback participants receive and the effect size reported for the corresponding study (Honorton & Ferrari, 1989; Steinkamp, Milton, & Morris, 1998). Tart (1977, 2001; Tart, Palmer, & Redington, 1979) has also argued for the inclusion of immediate feedback in parapsychological experiments. His claim revolves around a learning paradigm, in which some form of feedback on performance is almost always provided as reinforcement immediately after a behavioural response. The elimination or absence of feedback, in turn, is commonly used to extinguish a learned behavioural response to a given stimulus.

On the topic of feedback, the stimuli used in contingent tasks are also worthy of attention. In the Luke and colleagues' paradigm, participants in the negative reward condition had to take part in a boring number-vigilance task. In their replication attempt, Hitchman et al. (2012) observed that, despite most participants indicating that this task was relatively unpleasant, some actually claimed to enjoy the task, and it was felt that the emotive intensity of this negative reward could be enhanced. To ease the transition to a trial-by-trial feedback mechanism that would necessitate a larger number of incidences of contingent outcomes, it was also important to reduce the length of time required for each instance. Consequently, unpleasant, scary, or gruesome images were selected from the IAPS set (Lang & Greenwald, 1993) for use in negative reward conditions. Moreover, Hitchman et al. (2012) speculated that the primary function of psi-mediated instrumental responses may be to avoid negative outcomes. It was therefore thought that the use of strongly negative, gruesome, or scary images provided on a trial-by-trial basis may lead to a stronger aversion to the negative reward conditions, and hence a greater tendency for participants to exhibit PMIR during each trial.

The overall aim of the present study was to refine what has so far been a promising experimental protocol whilst exploring the roles of intentionality and lability in precognitive performance. It was predicted that participants would score more hits than would be expected by chance on both versions of the psi task, although better performance was expected during the nonintentional version. Participants' lability scores were also expected to be related to their precognitive performance. Furthermore, consistent with the Hitchman et al. (2012) study, hypotheses regarding the relationship between participants' psi scores and their beliefs about psi and their own luck were tested. Finally, it was predicted that precognition scores would be positively related to participants' scores on Bem's (2003, 2011) emotional reactivity items.

Method

Design

A repeated-measures, quasiexperimental design was employed in which participants completed both a 10-trial nonintentional precognition task and a 10-trial intentional precognition task in sequential order. In both cases, the dependent variable was the number of direct hits they scored on the task, when the mean chance expectation (MCE) was five hits for each participant. A trial-by-trial reward manipulation was utilised such that each time participants scored a hit ($p = .5$), they were administered a positive reward of seeing a pleasant image whereas each time they scored a miss (p = .5) they were given a negative reward of seeing a negative image. For correlational purposes, questionnaire measures were used to collect individual difference data for the independent variables of paranormal belief (sheep-goat), a composite lability measure, and an emotional reactivity measure.

Participants

Twenty male and 30 female participants were recruited by opportunity sampling from friends, colleagues, associates, students at the University of Northampton, interested members of the public, and members of local hobby groups. Although two participants did not disclose their age, the mean age of the remaining participants was 27.21 years (SD = 8.72). Participants were invited to take part in "a psychological investigation of possible psychic ability and how it relates to an individual's personality and beliefs." No incentives were offered in exchange for participation and all participants were informed that they may see unpleasant images before giving their final consent to take part in the study.

Individual Difference Measures

Demographic questionnaire. This 2-item questionnaire asked about participants' age and gender.

Sheep-goat belief questionnaire. This 5-item questionnaire contained four questions corresponding to different aspects of the belief in psi variable as specified by Palmer (1972) in addition to a fifth item concerning whether or not participants believed their luck could influence the outcome of the psi task. Each item is scored on a true/false basis, yielding a total belief score that can range from 0 to 4.

Openness to experience (OE) scale. This 20-item questionnaire addresses an individual's openness to new experiences (Goldberg, 1999). Participants respond to items such as "Believe in the importance of art" and "Have a rich vocabulary" by indicating the extent to which each statement is an accurate description of themselves. Each item is rated on a 5-point Likert scale from "very inaccurate" to "very accurate," yielding a score which can range from 0 to 80. Coefficient alphas for subscales range from .77 to .86 (Goldberg, 1999), and these scores have been found to correlate with scores on the equivalent scale of the NEO personality inventory (r = .56; Gow, Whiteman, Pattie, & Deary, 2005).

Creative Cognition Inventory (CCI). This 29-item questionnaire addresses the use of different cognitive styles in the creative process, considering factors such as heightened internal awareness, intuition, and playfulness (Holt, 2002). Questions are categorised according to linear (4 items) and nonlinear (25 items) subscales. The linear scale relates to logical analysis, planning, and careful selection of ideas; it has acceptable internal consistency (α = .72). The nonlinear scale relates to paying attention to internal states, playful cognition, ideas arising in states along the dream-wake continuum, and a sense of ideas coming from "something other"; it also has acceptable internal consistency (α = .92). Both scales were found to have adequate construct, convergent, and discriminant validity (Holt, 2007). Respondents indicate the extent to which experiences such as "Trusting hunches or instincts" and "Paying attention to visual imagery" are important to their being creative. Items are scored on 5-point Likert scales from "not at all important" to "extremely important," yielding a total score that can range from 4 to 20 for the linear subscale and 25 to 125 for the nonlinear subscale.

Complex Partial Epileptic Signs (CPES) cluster of the Personal Philosophy Inventory (PPI). This 16-item questionnaire containing items which relate to experiences similar or analogous to those reported by patients with unusual activity in the temporal lobes, often achieved by means of direct electrical stimulation during surgery (Persinger & Makarec, 1987). Each item is scored on a yes/no basis, yielding a total score that can range from 0 to 16. Patients' responses to these items have been found to correlate significantly with measures of electroencephalographic activity localised to the temporal lobes (Makarec & Persinger, 1987). The CPES scale was found to have a satisfactory level of internal reliability and test-retest correlations of $r = .85$ to $r = .95$ after 10 days and $r = .60$ to $r = .70$ after 100 days (Persinger & Valliant, 1985).

Emotional Creativity Inventory (ECI). This 30-item scale contains items that address components of emotional creativity such as preparedness, novelty, effectiveness, and authenticity (Averill, 1999). Participants respond to statements such as "My emotional reactions are different and unique" on a scale from 1 ("strongly disagree") to 5 ("strongly agree"), yielding a total score that can range from 30 to 150. Only two of the items are reverse scored, but no evidence was found of response bias in test data (Averill, 1999). The inventory has been found to have a high level of internal reliability ($\alpha = .90$) and test-retest reliability ($r = .91$ after 3 months).

Mood lability items. Three items derived from a self-report mood affect scale developed by Akiskal et al. (1995) based on the frequent up-and- down fluctuations in mood observed between sufferers' episodes of mood disorders such as bipolar II disorder (Inter-episode mood lability; Kraepelin, 1921). Participants indicate the extent to which each statement reflects patterns in their mood on a 3-point scale from "Not at all" to "Very much so," yielding a total score that can range from 3 to 9. These items have been found to be associated with clinical diagnoses of bipolar II disorder and family history of mood disorders (Benazzi, 2004; Benazzi & Akiskal, 2005).

Emotional reactivity items. These two items address individuals' awareness of their emotional reactivity to violent, scary, or gruesome content in photographs, movies, and videos (Bem, 2003, 2011). Participants respond on a scale from 1 ("not at all intensely aware") to 5 ("very intensely aware"). Bem advises the use of mean scores for correlational analysis, which can range from 1 to 5.

Materials for Test Session

PMIR visual basic program. The software program used in the Hitchman et al. (2012) study was modified for this experiment by the first author (available by request). The program's code incorporates a number of procedural changes that enabled the experimental hypotheses to be addressed. As before, all questionnaires were integrated into the software and the program was completely automated such that participants could give their responses and complete the experimental tasks without the presence, aid, or intervention of the experimenter.

International Affective Picture System (IAPS). The program presents a large set of emotive colour photographs from the IAPS set, the contents of which span numerous semantic categories including awe, excitement, contentment, amusement, fear, sadness, disgust, and anger (Lang & Greenwald, 1993). The images have been rated by independent judges for their perceived valence, arousal, and dominance. After all erotic images were filtered from the set on ethical grounds, the images were sorted into three groups: (a) positive images consisting of the most positively valenced pictures (mean pleasantness rating > 6), (b) negative images consisting of the most negatively valenced pictures (mean pleasantness rating < 4), and (c) neutral images consisting of those pictures with a mean pleasantness score around the midpoint of the rating scale (4.5 < mean pleasantness rating < 5.5). From each group, images were then paired together into sets such that each pair satisfied the following criteria: (a) both images depicted content which, in the authors' opinion, were drawn from similar semantic categories (e.g., a picture of rabbits and a picture of puppies both represent animal content), (b) the mean pleasantness and arousal ratings were very closely matched (for all pairs, pleasantness ratings were within .5 of a unit and arousal ratings were within 1.5 units), (c) the sum of the standard deviations of pleasantness and arousal ratings did not exceed 3.5 units (implying that the majority of individuals have similar emotional responses to the images). Using this method, 20 pairs of neutral images were selected to be used as target and decoy images for both nonintentional and intentional precognition trials. Similarly, 10 pairs of positive and 10 pairs of negative images were isolated for use as positive and negative reward images for the nonintentional precognition trials. Finally, 10 single positive and 10 single negative images were selected from the remaining unused images for use as positive and negative reward images in the intentional precognition trials.

Procedure

Participants were greeted and briefed in a quiet room, where it was explained to them that the experiment involved a test of psychic ability, but the implicit, precognitive nature of the nonintentional task was not disclosed. After being given the opportunity to pose any questions, participants were left alone in a quiet area to operate the computer program on a laptop computer. The principal investigator waited in a nearby room and was available to help if participants needed any further assistance.

The program displayed a written introduction before collecting participants' informed consent and presenting them with digitised versions of the questionnaire battery. Subsequent to participants answering all of the questions, the program instructed them to proceed to the experimental tasks. The program then gave instructions for what had been described to them in the briefing as a "preparatory" image preference indication task. Participants were told they would be shown some pictures in pairs and asked to choose their preferred image from each set. As a rationale for this task, participants were speciously told that their selections would help the program to choose appropriate targets for a later task.

Participants then took part in 10 nonintentional precognition trials. During each trial, they were shown one of the neutral pairs of images and selected which one of the two images they most preferred. An example neutral image pair is presented in Figure 1. Please note that for contractual reasons, indicative images are displayed rather than genuine IAPS pictures in all figures. At the beginning of each trial, the position of the cursor was reset to the centre of the screen to avoid biasing participants towards selecting either of the images. The trial in which each specific pair of images occurred was randomised for each participant, and the place each of the two images from each pair appeared on the screen (left or right) was also randomised for each trial.

Figure 1. Example stimuli for the precognition task.

Unbeknown to the participants, this image preference task constituted a forced-choice, nonintentional precognition task, as each time they indicated their preferred image from the pair the computer immediately selected one of the images at random as the target. Directly after each trial, participants were administered a positive or negative reward for their performance by means of a secondary image preference task. If the participants' selection matched the computer's random selection, they entered the positive reward condition, and the secondary image preference task consisted of participants indicating their preferred image from one of the pairs of positive images (Figure 2). However, if the participants' selection didn't match the computer's random selection, they entered the negative reward condition, and the secondary image preference task consisted of participants indicating their preferred image from one of the pairs of negative images (Figure 3). No data were collected regarding participants' selections during the contingent reward tasks.

Figure 2. Example stimuli for the positive reward condition

Figure 3. Example stimuli for the negative reward condition

As per the Hitchman et al. (2012) study, randomisation of the on-screen image array positions and computer target selections was achieved using the random number generation function within VB.NET, which is seeded by the CPU timer. Goodness of fit tests were conducted on the experimental data to evaluate whether or not adequate randomisation had been achieved. The tests revealed there was no bias in the number of times each target was selected by the computer, $\chi^2(1, N = 1000) = 0.68$, $p = .41$), nor a left/right bias in the positioning of stimulus images on screen, $\chi^2(1, N = 1000) = .06$, $p = .80$.

After participants had received their positive or negative reward for the 10th nonintentional trial, the computer displayed a dialogue box that informed them they had completed the task and asked them to call back the experimenter. The experimenter informed the participants of the implicit, precognitive nature of the task they had just completed and answered any questions to ensure that this had been thoroughly understood. Here, it was ensured that participants had been unaware of the true nature of the task they had just completed, as a means to validate that they had not intentionally attempted to exhibit precognition within trials. Participants were then told that the final part of the experiment involved a very similar task. However, now they were aware of how the program functioned, rather than simply selecting their favourite of the neutral images, participants were asked to use their will and intent to try to predict (by whichever means they wished) which image they thought the computer would select, thus achieving as many of the positive contingent rewards as possible. Once they had confirmed their understanding of the instructions, participants were provided with a password that was required for them to proceed to the final task on the computer.

After participants entered the password, the computer reiterated the instructions for the final task, prompting them to intentionally attempt to select the images which they believed would lead them to the positive outcomes. The program once again proceeded through 10 precognitive trials by displaying pairs of closely matched neutral images in random order. As it was no longer necessary to maintain any secrecy, in the reward conditions, rather than giving participants with a choice between pairs of positive or negative images, a single positive or negative reward image was displayed on the screen for 3 s before automatically proceeding to the next trial.

Once participants had participated in and been positively or negatively rewarded for all 10 intentional precognition trials, the computer informed them that they had completed the experiment and prompted them to call back the experimenter. The experimenter then provided a full debrief. Typically, participants discussed their impressions of how well they thought they had performed in each version (nonintentional and intentional) of the task. Before leaving, all participants were asked not to discuss the nature of the experiment with other potential participants.

Ethics

The project was designed to adhere to the British Psychological Society's Code of Ethics and Conduct (BPS, 2009) and received ethical approval from the University of Northampton Research Ethics Committee. Participants were briefed prior to giving their informed consent as part of the program. All data were collected anonymously and participants were made aware of their right to withdraw from the experiment at any time without having to provide a reason. Importantly, participants were also forewarned in the briefing period that they may see negative, gruesome and scary images during the experiment, and it was also mandatory that they ticked a box indicating that this had been explained to them and they were happy to continue to take part in the study to proceed past the digital consent form.

Results

The total number of hits participants scored on the nonintentional and intentional versions of the psi task was recorded, along with their scores on the individual difference measures. One participant did not provide answers to a number of items of the nonlinear subscale of the Creative Cognition Inventory and two participants did not provide sufficient data in relation to the Emotional Creativity Inventory. Due to the large number of omitted responses for these participants, substitution of scores was not considered feasible and consequently those participants' data have been excluded from analyses where appropriate.

The primary hypotheses predicted that participants would select more target images during the nonintentional and intentional versions of the precognition task than would be expected by chance alone. Fifty participants each completed 10 nonintentional and 10 intentional precognition trials with an associated probability of correctly selecting the target image of .50. Thus, with a total of 500 trials for each version of the task, the MCE was 250 hits. For the nonintentional precognition task, the actual number of hits was 247, with a mean hit rate per participant of 4.94 hits (SD = 1.17). As participants scored marginally fewer hits than the mean chance expectation (MCE = 5.00 hits), the experimental hypothesis was not supported. The representative effect size by t test was r = -.02 compared to r = .16 observed by Hitchman et al. (2012). In the intentional version of the task, participants accumulated a total of 245 hits, marginally lower than the MCE of 250. The mean hit rate per participant was 4.90 hits (SD = 1.42). As participants scored slightly fewer hits than the MCE, the experimental hypothesis was not supported. The representative effect size was r = -.03. On average, participants performed marginally better at the nonintentional version of the task (M = 4.94 hits, SD = 1.17 vs. M = 4.90 hits, SD = 1.42), but the difference in performance across the two conditions was not statistically significant, $t(49) = 0.16$, $p = .44$, one-tailed. Furthermore, there was only a very slight positive correlation between intentional and nonintentional precognition scores, $r(48) = .07$, $p = .31$, one-tailed.

The next set of hypotheses concerned relationships between psi task performance and individual difference measures related to lability. For transparency in interpreting the correlations between psi task scores and individual difference scores reported below, Table 1 presents a correlation matrix of these individual difference measures to highlight where there may be shared variance.

Table 1

Pearson Correlations and Two-Tailed Significance Values Between Predictors of Psi Task Performance (N = 50)

	Sheep-Goat	ECI	Mood Affect	OE	CCI Linear	CCI Nonlinear	PPI
ECI	.34*						
Mood Affect	.07	.04					
OE	.15	.49***	-.03				
CCI Linear	-.16	.05	-.02	-.10			
CCI Nonlinear	.37**	.63*****	.06	.53****	.01		
PPI	.51****	.38**	.05	.42**	-.24	.50***	
ER	.09	.34*	.04	.26	-.09	.17	.04

Note. ECI: Emotional Creativity Inventory; OE: Openness to Experience; CCI: Creative Cognition Inventory; PPI: Personal Philosophy Inventory; ER: Emotional Reactivity
*p < .05. **p < .01. ***p < .001. ****p < .0001. *****p < .00001.

It was predicted that participants' performance on the (a) nonintentional and (b) intentional versions of the precognition task would be positively correlated with their scores on a composite measure of lability. Scores on the constituent elements of lability (Emotional Creativity, Mood Affect, Openness to Experience, Creative Cognition and Complex Partial Epileptic Signs) were amalgamated into a single score. Table 2 shows that composite lability scores were positively related to participants' performance on the nonintentional version of the precognition task, although the relationship was not statistically significant, $r(45) = .17$, $p = .13$, one-tailed. Conversely, contrary to the alternative hypothesis, composite lability scores were found to be negatively related to participants' performance on the intentional version of the precognition task, $r(45) = -.13$. A Steiger calculation (Clark-Carter, 2010) revealed that the difference between the two correlations is not significant, $t(44) = 1.39$, $p = .09$, one-tailed. Furthermore, none of the constituent elements of the composite lability scale were found to be significantly related to either intentional or nonintentional precognition scores.

Table 2
Pearson Correlations Between Psi Task Scores and Predicted Covariates (Steiger One-Tailed Significance)

	N	Nonintentional	Intentional	Steiger's *t*
Sheep-goat	50	.07 (.31)	-.06 (N/A)	0.68 (.25)
ECI	48	.06 (.34)	-.13 (N/A)	0.91 (.18)
Mood Affect	50	.03 (.42)	-.10 (N/A)	0.68 (.25)
OE	50	.10 (.26)	-.12 (N/A)	1.10 (.14)
CCI Linear	50	.18 (.11)	.22 (.06)	0.24 (.41)
CCI Nonlinear	49	.13 (.18)	-.08 (N/A)	1.03 (.15)
PPI	50	.18 (.11)	-.20 (N/A)	2.01* (.03)
Composite Lability	47	.17 (.13)	-.13 (N/A)	1.39 (.09)
Emotional Reactivity	50	.29* (.02)	-.04 (N/A)	1.68 (.05)

Note. ECI: Emotional Creativity Inventory; OE: Openness to Experience; CCI: Creative Cognition Inventory; PPI: Personal Philosophy Inventory
*$p < .05$.

It was also predicted that emotionally reactive participants would perform better at both versions of the precognition task. A mean score of the two emotional reactivity questions was calculated and correlated against participants' precognition scores. For nonintentional trials, there was a significant positive correlation between participants' precognitive performance and their mean emotional reactivity scores, $r(48) = .29$, $p = .02$, one-tailed. Contrary to the alternative hypothesis, for intentional trials there was a small, negative, correlation between the two variables, $r(48) = -.04$. A Steiger calculation revealed that the difference between the two correlations was on the threshold of statistical significance, $t(47) = 1.68$, $p = .05$, one-tailed.

Turning to the belief measures, consistent with the Hitchman et al. (2012) study, it was predicted that performance on the (a) nonintentional and (b) intentional versions of the precognition task would be positively correlated with participants' belief in psi (as measured by criteria I-IV of the Sheep-Goat questionnaire). Table 2 shows a very small positive correlation between belief in psi and performance on the nonintentional version of the precognition task and, unexpectedly, a very small negative correlation between belief in psi and performance on the intentional version of the precognition task, nonintentional: $r(48) = .07$, $p = .31$, one-tailed; intentional: $r(48) = -.06$.

Also based on the Hitchman et al. (2012) study, it was hypothesised that participants who believed they could use their psi to affect the outcome of the experiment would perform better at the (a) nonintentional and (b) intentional versions of the precognition task. Participants responded true or false to the statement "I believe that my psychic ability can affect the outcome of this experiment" and were grouped according to their response. The majority of participants ($n = 40$) indicated they were goats (i.e., their psychic ability couldn't influence the outcome of the experiment) compared with just 10 sheep. Sheep performed marginally better at the nonintentional version of the precognition task than goats (sheep mean precognition task score = 5.30, $SD = 0.82$; goat mean precognition task score = 4.85, $SD = 1.23$), but the difference was not significant, $t(48) = 1.09$, $p = .14$, one-tailed. Contrary to the alternative hypothesis, in the intentional version of the task, goats outperformed sheep (sheep mean precognition task score = 4.60, $SD = 1.35$; goat mean precognition task score = 4.98, $SD = 1.44$).

A similar hypothesis predicted that participants who believed that their luck could influence the outcome of the test would achieve a higher score on the (a) nonintentional and (b) intentional versions of the precognition task. For nonintentional trials, the 30 "luck sheep" (60%), who indicated that they believed their luck could affect the outcome of the experiment, scored nonsignificantly higher on the psi task than the 20 "luck goats" (40%), who indicated to the contrary (luck sheep mean psi score = 5.13, $SD = 1.31$; luck goat mean psi score = 4.65, $SD = .88$; $t(48) = 1.45$, $p = .08$, one-tailed. For intentional trials, luck sheep also outperformed luck goats (luck sheep mean psi score = 5.20, $SD = 1.32$; luck goat mean psi score = 4.45, $SD = 1.47$). The result of an independent samples t test indicated that luck sheep performed significantly better than the luck goats, $t(48) = 1.88$, $p = .03$, one-tailed.

As noted by Hitchman et al. (2012), the parametric Pearson correlation test may not be valid for scales with a score range of less than 20 points (Clark-Carter, 2010). This applies to the Sheep-Goat, Mood Affect, Personal Philosophy Inventory and Emotional Reactivity measures. Whilst Pearson correlations are reported above for consistency and comparison with other correlations, Spearman nonparametric correlations were also calculated. The results indicated the same pattern: Sheep-oat: nonintentional, $r_s = .03$ $p = .42$, one-tailed; intentional, $r_s = -.07$. For Mood Affect: nonintentional, $r_s = .03$, $p = .42$, one-tailed; intentional, $r_s = -.08$; Personal Philosophy Inventory: nonintentional, $r_s = .16$, $p = .14$, one-tailed; intentional, $r_s = -.22$; Emotional Reactivity: nonintentional, $r_s = .36$; $p = .005$, one-tailed, intentional; $r_s = -.10$.

Post-Hoc Analysis of Participants' Image Preference Biases

The present study made use of closely matched pairs of authentic images as targets in the precognition task as opposed to the sets of four fractal images used in previous studies. It was therefore important to assess whether participants exhibited any systematic biases towards either image from each pair. Table 3 presents the number of times each image from each set was selected. The results of chi-square analyses indicated that participants appeared to exhibit a systematic preference for image A in set 8, χ^2 (1, $N = 50$) = 3.92, $p = .05$ and image B in set 11 χ^2 (1, $N = 50$) = 5.12, $p = .02$. Hitchman et al. (2012) reported that participants showed similar biases within 2 out of the 15 sets of fractal target images in their study. Consequently, it would appear that the use of authentic images in place of fractal target images did not bring about a tangible increase in the extent of participants' preferences for specific images within each set.

Table 3
Chi Square Analyses of Participants' Preferences for a Specific Image from Each Target Pair

Image set	Image A selected	Image B selected	χ^2	*p*
1	31	19	2.88	.09
2	20	30	2.00	.16
3	29	21	1.28	.26
4	19	31	2.88	.09
5	22	28	0.72	.40
6	24	26	0.08	.78
7	26	24	0.08	.78
8	32	18	3.92	**.05**
9	30	20	2.00	.16
10	24	26	0.08	.78
11	17	33	5.12	**.02**
12	31	19	2.88	.09
13	25	25	0.00	1.00
14	29	21	1.28	.26
15	30	20	2.00	.16
16	19	31	2.88	.09
17	21	29	1.28	.26
18	30	20	2.00	.16
19	31	19	2.88	.09
20	26	24	0.08	.78

Post-Hoc Analysis of Practice/Decline Effects

Given that by necessity in this repeated measures design participants always took part in the nonintentional precognition task before completing the intentional precognition task, concerns were raised over the potential influence of practice and/or decline effects. In order to assess this, the total number of hits achieved by all participants for each trial was considered. Figure 4 indicates that there were no clear patterns in performance across either nonintentional trials (T1-T10) or intentional trials (T11-T20). In support of this, chi-square analysis indicated there were no significant differences between overall performance in each trial, χ^2 (19, N = 50) = 8.65, p = .98. Therefore, no evidence of either practice or decline effects was observed across the experiment.

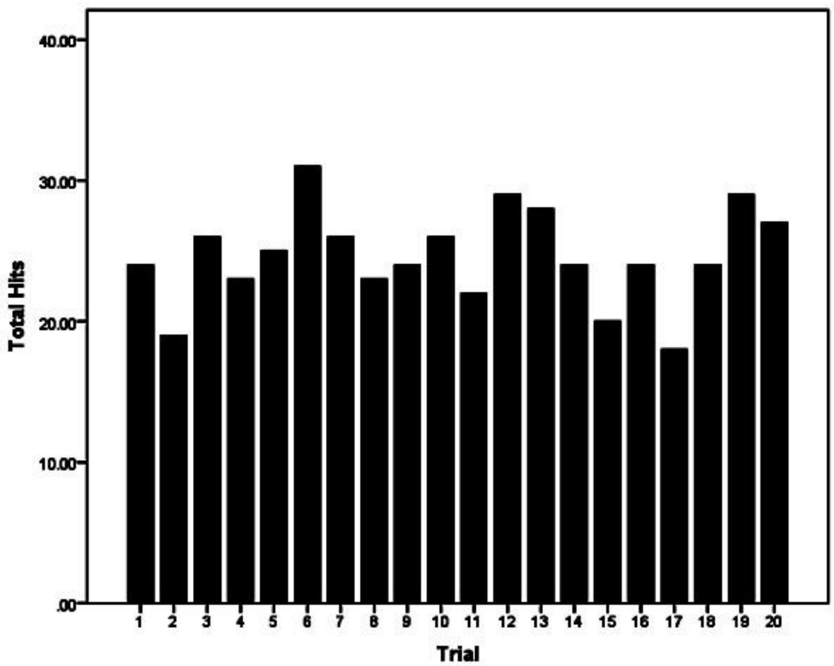

Figure 4: Total hits achieved by all participants for each trial.

Consistent with the Hitchman et al. (2012) study, multiple statistical tests have been conducted without a correction applied to the alpha levels for multiple analyses. Milton and Wiseman (1997) have noted that the standard Bonferroni adjustment should be considered conservative, whilst Abdi (2007) claims the Bonferroni correction is not appropriate when the inferential tests conducted are not entirely independent. Readers are advised that due to the exploratory nature of this study, all significant results reported within this manuscript are claimed as tentative pending replication, especially as the chance of a type 1 error is increased as a consequence of the multiple analyses carried out.

Discussion

The main experimental hypotheses of this study concerned a comparison of performance at equivalent nonintentional and intentional precognition tasks and the relationship between precognition and lability scores. Within Stanford's specification of the PMIR model, it is claimed that knowledge of need-relevant circumstances and an intention to fulfil such needs may play an inhibitory role in the psi-mediated instrumental response process. It was therefore expected that performance on the nonintentional precognition task would be higher than on the intentional precognition task. Overall, participants selected slightly fewer of the target images than the MCE in both nonintentional and intentional versions of the precognition task, providing no evidence of either intentional or nonintentional psi. Furthermore, although participants performed slightly better in the nonintentional version of the task as hypothesised, there was not a statistically significant difference between scores. Meanwhile, participants' lability scores were found to be unrelated to both their intentional and nonintentional precognitive performance.

Although the results of the Hitchman et al. (2012) study in relation to the main psi effect were in the predicted direction, they were nevertheless nonsignificant. The findings of the present study therefore represent a further decline in tacit precognition scores to below-chance levels and a second failure to replicate the significant effects demonstrated by Luke and associates using a largely similar method. Colborn (2004) has reviewed a multitude of factors that may account for general patterns of declining results across parapsychological paradigms. Amongst these, there are several potential explanations that may account for these dissimilarities in results obtained by different researchers, including false positives (Type I errors), decline effects, design modifications, and experimenter effects. In the case of the former, it may simply be a chance occurrence that Luke and associates were able to achieve psi indicative results that are nevertheless spurious and not reflective of a genuine and robust effect. However, this interpretation is questionable given that the results were replicated across a series of four studies, and are consistent with a much larger database of similar psi-indicative research in the forced-choice psi paradigm (e.g., Honorton & Ferrari, 1989).

With respect to design modifications, one key area in which this study differed from previous studies was in the implementation of trial-by-trial feedback. Data from several meta-analyses of forced-choice intentional psi studies indicate that the immediacy of feedback in relation to the psi task is a key variable in determining the size of effects (Honorton & Ferrari, 1989; Steinkamp, 2005; Steinkamp et al., 1998). The use of trial-by-trial feedback also helped to overcome a potential issue that the fundamental need being fulfilled in the psi task may be the avoidance of punishment, rather than the seeking of the highest reward conditions. With trial-by-trial feedback, it is necessary for participants to score a hit in each and every trial to avoid being negatively rewarded, whereas the feedback system in previous studies only required participants to score above chance across the entire run of trials to escape punishment. Despite the conceptual advantages of this feedback mechanism, participants, on average, failed to outperform MCE in the present study, whereas above-MCE results have been reported in each of the studies of this type employing end-of-run feedback. It may be, then, that in the context of this type of experiment, a series of smaller rewards carry less weight than a single reward of a longer duration.

This study also differed in terms of the task used in the negative reward condition. In the previous study, participants who underperformed the mean chance expectation over 15 trials took part in a boring number vigilance task. In the present study, however, participants were shown a negative image from the IAPS picture set which contained violent, gruesome, or scary content. This was intended to enhance the emotive impact of the negative outcome and hence increase participants' aversion to the negative reward condition. This rationale was supported by the finding that a significant, positive relationship was observed between participants' performance on nonintentional trials and their mean scores on Bem's (2003, 2011) emotional reactivity items.

However, the use of trial-by-trial feedback ensured that it was very unlikely ($p < .001$) that participants could entirely avoid negative rewards, whereas in previous end-of-run scenarios, there was a 50% chance of avoiding the negative contingent task. It is therefore possible that the presence of unpleasant stimuli throughout the experiment potentially induced a general state of anxiety or a psi equivalent of learned helplessness in some participants, which may have manifested itself within trials. As a result, rather than holding a consistent state of openness, being sensitive to extrasensory stimuli and responding to them accordingly, participants may instead have maintained feelings of apprehension and defensiveness. It is also a possibility that certain positive or negative reward images could have primed participants' selections of any subsequent neutral target images. Furthermore, a number of participants indicated a level of curiosity towards the negative images, suggesting that some may have been more attracted to the negative reward condition than the positive reward condition. In hindsight then, the use of emotive images as feedback may be better suited to an end-of-run feedback system to avoid building and reinforcing a state of anxiety rather than openness and potentially priming subsequent decision making.

With respect to the target images themselves, on the basis of participant and reviewer feedback in response to the Hitchman et al. (2012) study, sets of four fractal patterns were replaced by pairs of authentic images in the present study. This was intended to enhance the ecological validity of the precognition task. However, in a similar manner to the Hitchman et al. (2012) study, it was found that participants exhibited a systematic bias towards one of the images in two of the target sets. Although in the case of both studies target images were selected on the basis of being closely matched in terms of their content and valence and arousal ratings, in any set of images which differ tangibly in appearance, there remains an opportunity for a range of preferential biases (e.g. colour, gender) to influence participants' selections. Although it could come at the cost of psi discriminability, it is recommended that future studies attempt to further increase the normalisation of image sets, such as by using pairs of mirrored images, as in Bem (2011).

On the topic of ecological validity, it is worthwhile to consider a further design element at this stage. In the present study, participants took part in a total of 20 equivalent trials. Much of the early PMIR work (e.g. Stanford & Thompson, 1973) typically involved a single opportunity for participants to use psi to achieve a need-relevant goal, rather than a series of repetitious trials. Although many of the case reports from which Stanford developed the PMIR model appeared to involve only one opportunity for the exhibition of psi to achieve a favourable outcome (Stanford, 1974), repetition per se is not necessarily atypical. For example, lots of small delays in a bookshop could just as easily lead to a serendipitous meeting as a single delay of a longer duration. Nevertheless, it is recommended that future studies pay very careful attention to the ecological validity of psi tasks to ensure that they reflect the ways in which psi is assumed to occur in everyday life situations.

Turning to the role of intentionality in this paradigm, this study failed to yield evidence of the potentially inhibitory influence of need-relevant information or any conscious cognitions in relation to such information in psi-mediated scenarios. According to Stanford (1990), all that is necessary for an adaptive psi-mediated outcome to occur is a behaviour, and any consciously generated thinking, cognitive constraints or need-relevant information in relation to such a behaviour could inhibit the potential for it to be mediated by psi. In the present study, participants were foretold in their briefing that a psi task (the intentional psi task) would follow what was described as a "preparatory" image preference task (the nonintentional psi task). Participants performed similarly when offered a minimal amount of informational cognitive priming (as in the nonintentional task) and when given full disclosure regarding the nature of the task (as in the intentional task). Although no measure of cognitive activity was employed to objectively assess whether or not there were differences in the conscious cognitions of participants in the nonintentional and intentional conditions, it is unlikely that they would have felt the need to doubt or disobey the instructions they were provided. Consequently, it is reasonable to assume that in the nonintentional condition participants were simply indicating their image preferences, whereas in the intentional condition they were actively engaged in trying to foretell which image would be randomly selected by the computer. According to the authors' interpretation, the cognitive constraints which could inhibit psi were defined rather broadly in the specification of the PMIR model, so it is difficult to assess whether or not the different versions of the task would have resulted in participants engaging in different cognitive activities to an extent which Stanford believed may have a tangible impact on the psi process. Furthermore, in the absence of objective empirical data in relation to participants' cognitive processes during the tasks, it is not possible to discount the possibility that the limited information participants were given about the eventual intentional psi task may have led them to experience conscious cognitions that influenced either the speed of their decisions or the decisions themselves within the nonintentional psi task. Whilst greater efforts could be made to further reduce the cognitive priming given to participants in relation to the implicit psi task in future studies, we must still face the issue that if we accept the psi hypothesis as valid, it would not be possible to entirely avoid cognitive priming, as information in relation to tacit tasks could be available by extrasensory means.

It is also important to give mention to a design compromise in the present study. Given resource and time constraints, a repeated measures design was employed in order to meet sample requirements. Ideally, conditions in repeated measures designs should be counterbalanced in order to compensate for the potential for range effects such as practice, sensitisation and carry over effects from confounding results (Clark-Carter, 2010; Greenwald, 1976; Poulton, 1973). However, given that the nonintentional version of the precognition task relied on the naivety of the participants, it was entirely necessary for participants to take part in the nonintentional version of the task prior to receiving the briefing for the intentional version of the task. As a result, it is possible that some participants may have exhibited a greater level of performance at the intentional version of the task owing to them having practised an equivalent task in the previous condition. Conversely, some participants may have exhibited a diminished level of performance in the intentional trials owing to a decline effect, boredom, desensitisation to the reward stimuli or any other performance-based carry over effects associated with having performed a similar task in the previous condition. Due to this task order confound, any conclusions based on these results cannot be relied upon with confidence. Although chi square analysis found no evidence of consistent improvements or declines in performance across the experiment, authors typically recommended against the use of within subjects designs when the juxtaposition of conditions in not the main factor of interest (Greenwald, 1976; Poulton, 1973). The potential for these effects to manifest themselves within the data could only be eliminated by randomly allocating participants to take part exclusively in one of the two conditions in a between-subjects design.

The secondary hypotheses in this study concerned the performance of individuals at the nonintentional and intentional versions of the precognition task in relation to individual difference measures. Stanford (1990) had proposed rigidity in thought and behaviour as one of the principal inhibitory factors in the PMIR model. However, no support was found for the hypothesised effect of lability, with participants' scores on the composite measure not being found to correlate significantly with their performance at nonintentional or intentional precognition trials. Moreover, none of the constituent elements of the lability scale were found to covary to a significant extent with precognition scores. Focusing purely on effect sizes rather than statistical significance does not provide much more encouragement: all correlations were below r = .2 in absolute size. Similarly, no relationships were found between precognition task performance and participants' paranormal beliefs.

However, a significant, positive relationship was observed between participants' precognitive performance and their mean scores on Bem's (2011) Emotional Reactivity items, but only for nonintentional trials. Given the transition to using more emotionally potent images in the negative reward condition in this study, it was deemed particularly pertinent to have a gauge of whether individuals who were more reactive to negative emotive content would be more aversive to these images, and hence avoid the negative reward condition more frequently. As this relationship was only found to be significant for the nonintentional condition, the condition more similar to Bem's (2011) precognitive habituation task in which a similar effect has been observed, it was interesting to note that a Steiger calculation revealed the difference between the correlations for international and nonintentional trials approached a statistically significant level. This may suggest that participants' emotional reactivity does not interact as strongly with their tacit psi performance when they are consciously aware of the need to use some form of precognition.

As was noted by Hitchman et al. (2012), attempting to identify predictors of performance in a psi task in an experiment where no overall psi effect has been observed is not straightforward, as it is unclear whether psi phenomena were entirely absent from the experimental scenario, or simply if the majority of participants failed to demonstrate this ability. Indeed, Palmer (2009) has echoed these concerns by bemoaning the unreliability of psi performance and the effect of this on attempts to assess its covariates. Unless a relatively consistent psi effect can be identified, attempting to assess the roles of the awareness of the need to use psi, the intent to use psi and individual difference correlates of psi will remain problematic to achieve. What's more, a further concern with assessing trait-based measures as predictors of experimental variables is that that having a particular trait does not necessarily mean that the trait will be expressed under all circumstances. Consequently, future studies may wish to attempt to validate that any measure of lability used for a similar purpose is predictive of participants' propensity to exhibit a relevant labile state within the context of the experimental task, as well as ensure that the task is of such a nature that participants in that labile state are more likely to achieve a successful outcome.

To confound this issue, it is important to note that the reliability and validity of psychometric measures are typically assessed in isolation (i.e., when not administered amidst multiple other tests). However, Council's (1993) paper on context effects (CEs) highlights that correlations between psychological tests can vary considerably depending on whether they have been administered in the same testing session. The distortion of outcomes due to CEs has not been studied extensively, but it is important to keep in mind that the use of multiple measures within the same session can have unknown consequences and potentially threaten the construct validity of tests.

In addition, administering person-based measures before a test of psi can potentially result in a range of reactivity-related issues including, but not limited to, demand characteristics, priming, and rumination. Future researchers may seek to keep the measurement of predictor variables more distant from the experimental situation, perhaps in a separate controlled testing session. Furthermore, it is worthwhile to note that although participants received standardised instructions before beginning the experimental tasks, there were some minor differences in the information participants were provided with prior to attending their experimental sessions. Most importantly, for some participants who took part following the recommendations of their associates, it was not possible to inform them that the study involved the likelihood of seeing unpleasant images before their arrival. Consequently, it is possible that their decision to consent to take part in the experiment may have been influenced by the fact that they had already made a considerable effort to attend the testing session prior to being fully informed. As well as being an ethical concern which should be avoided in future, it is also possible that this and other subtle differences in the information participants were provided prior to attending the testing session may have affected their orientation to the experimental tasks and their subsequent performance.

Nevertheless, for the second time, a significant difference was found between the psi task performance of luck sheep, those who believed they could use their luck to influence the outcome of the experiment and luck goats, those who didn't believe they could do so. This finding, then, further strengthens the notion that a person's expectation in their ability to use luck in a particular situation may play a greater role in their success as opposed to the specific ways in which they conceive of luck (see Hitchman et al., 2012). It should be considered worthwhile to include the luck sheep-goat variable in subsequent studies to assess whether it is able to withstand the test of time and prove to be a robust and reliable effect.

Overall, this study has advanced the nonintentional precognition paradigm in several ways and, in turn, has raised a number of additional questions. Firstly, this experiment addressed the difference between nonintentional and intentional psi tasks, particularly in relation to the cognitive activities of the participants engaging in the tasks. No difference was found between performance at either version of the precognition task, raising doubt over the PMIR model's assertion that cognitive priming and focused intent can diminish the potential for psi-mediated instrumental responses to be executed. However, given the potential confound of task order and the fact that no overall evidence of psi was found within either version of the task, such doubts should only be cast cautiously. Concurrently, the first signs of a general pattern of declining results in this paradigm were observed. A number of potential explanations for such findings have been considered and should be continually monitored as the paradigm develops. Looking ahead, experimenter effects may be one of the most interesting avenues to explore in future studies in an effort to account for differential results across contrasting experimental teams. Nevertheless, a meta-analysis considering the six Luke and colleagues and Hitchman and colleagues studies conducted using this paradigm to date suggests that the overall paradigm still presents significant evidence of tacit psi, with a Stouffer Z of 3.75, $p = .00008$, mean effect size $r = .19$.

This study has also contributed to the consideration of the factors that may aid or hinder the instigation of psi-mediated instrumental responses, with a particular focus on the role of lability. Overall, little evidence was found that lability or any of its constituent elements had a bearing on participants' precognitive performance. Conclusions in relation to the effect of individual difference covariates of psi are clearly restricted in a study devoid of any evidence of psi per se. Nevertheless, effects of luck beliefs and emotional reactivity were observed, which should be considered as worthwhile variables to include in studies henceforth. Going forward, researchers may wish to turn their attention towards developing more reliable performance-based measures of the other individual difference covariates that are predicted to influence the PMIR process, particularly latent inhibition. In doing so, every effort should be made to tailor trait-based measures to the context in which they are expected to be expressed and to minimise the potential for measurements to impact the assessment of other experimental variables.

References

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp.103–107). Thousand Oaks, CA: Sage.

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

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

Bem, D. J. (2003). Precognitive habituation: Replicable evidence for a process of anomalous cognition. *Proceedings of Presented Papers: The Parapsychological Association 46th Annual Convention*, 6-20.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425.

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

Benazzi, F. (2004). Validating Angst's ups & downs personality trait as a new marker of bipolar II disorder. *European Archives of Psychiatry and Clinical Neuroscience, 254*, 48–54.

Benazzi, F., & Akiskal, H. S. (2005). A downscaled practical measure of mood lability as a screening tool for bipolar II. *Journal of Affective Disorders, 84*, 225–232.

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

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

Braud, W. (2002). Psi-favorable conditions. In V. G. Rammohan (Ed.), *New frontiers of human science: A Festschrift for K. Ramakrishna Rao* (pp. 95–118), Jefferson, NC: McFarland.

Braud, W. G., & Schlitz, M. J. (1983). Psychokinetic influence on electrodermal activity. *Journal of Parapsychology, 47*, 95–119.

Braud, W. G., Shafer, D., & Mulgrew, J. (1983). Psi functioning and assessed cognitive lability. *Journal of the American Society for Psychical Research, 77*, 193–208.

BPS (2009). *Code of ethics and conduct: Guidance published by the Ethics Committee of the British Psychological Society*. Leicester, England: British Psychological Society.

Clark-Carter, D. (2010). *Quantitative psychological research: The complete student's companion*. New York, NY: Psychology Press.

Colborn, M. (2004). The decline effect in spontaneous and experimental psychical research. *Journal of the Society for Psychical Research, 71*, 1–21.

Costa, P., & McCrae, R. (1992). *Professional manual: Revised NEO Personality Inventory (NEO PI--R) and NEO Five-factor Inventory (FFI)*. Odessa, FL: Psychological Assessment Resources.

Council, J. R. (1993). Context effects in personality research. *Current Directions in Psychological Science, 2*, 31–34.

Drennan, S., Roe, C., & Broughton, R. (2011). Lability, paranormal beliefs and psychokinetic experiences: Development of the Lability Scale using an online survey-based study. *Abstracts of Presented Papers: The Parapsychological Association 54th Annual Convention*, 25-26.

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

Gow, A. J., Whiteman, M. C., Pattie, A., & Deary I. J. (2005). Goldberg's IPIP Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences, 39*, 317–329.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, *83*, 314–320.

Hitchman, G. A., Roe, C. A., & Sherwood, S. J. (2012). A re-examination of nonintentional precognition with openness to experience, creativity, psi beliefs and luck beliefs as predictors of success. *Journal of Parapsychology, 76*, 109–145.

Honorton, C., & Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology*, *53*, 281–308.

Holt, N. J. (2002). *Creativity and boundary permeability*. Unpublished manuscript, University of Northampton, UK.

Holt, N. (2007). *Creativity, states of consciousness and anomalous cognition: The role of epistemological flexibility in the creative process*. Unpublished doctoral thesis, University of Northampton, England.

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

Holt, N., Simmonds-Moore, C., & Moore, S. (2007). Psi and cognitive disinhibition: Exploring the filters of consciousness hypothesis. *Proceedings of Presented Papers: The Parapsychological Association 50th Annual Convention*, 192–194.

Kraepelin E. (1921). *Manic-depressive insanity and paranoia*. Edinburgh, Scotland: E & S Livingstone.

Lang, P. J., & Greenwald, M. K. (1993). *International affective picture system standardisation procedure and results for affective judgement: Technical reports 1A-1C*: University of Florida, Centre for Research in Psychophysiology, Gainesville, FL.

Lubow, R. (1989). *Latent Inhibition and conditioned attention theory*. Cambridge, England: Cambridge University Press.

Luke, D. P., Delanoy, D., & Sherwood. S. J. (2008). Psi may look like luck: Perceived luckiness and beliefs about luck in relation to precognition. *Journal of the Society for Psychical Research ,72*, 193–207.

Luke, D. P., & Morin, S. (2009, September). *Luck beliefs, PMIR, psi and the sheep-goat effect: A replication*. Paper presented at the meeting of the Society for Psychical Research, Nottingham, England.

Luke, D. P., Roe, C., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology, 72*, 133–154.

Makarec, K., & Persinger, M. A. (1987). Electroencephalographic correlates of temporal lobe signs and imaginings. *Perceptual and Motor Skills, 64*, 1124–1126.

Milton, J., & Wiseman, R. (1997). *Guidelines for extrasensory perception research*. Hatfield, England: University of Hertfordshire Press.

Palmer, J. (1972). Scoring in ESP tests as a function of ESP. Part 2: Beyond the sheep-goat effect. *Journal of the American Society for Psychical Research, 66*, 1–26.

Palmer, J. (2009). Winning over the scientific mainstream, *Journal of Parapsychology, 73*, 3–8.

Persinger, M. A. (1989). Geophysical variables and behavior: Predicting the details of visitor experiences and the personality of experients: The temporal lobe factor. *Perceptual and Motor Skills, 68*, 55–65.

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

Persinger, M. A., & Valliant, P. M. (1985). Temporal lobe signs and reports of subjective paranormal experiences in a normal population: A replication. *Perceptual and Motor Skills, 60*, 903–909.

Poulton, E. C. (1973). Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin, 80*, 113–121.

Rao, K. R., & Davis, J. W. (1978). The differential effect and experimenter effects in intentional and nonintentional psi tests. *Journal of Parapsychology, 42*, 1–19.

Roe, C. A., & Holt, N. J. (2006). The effects of strategy (“willing” versus absorption) and feedback (immediate versus delayed) on PK performance. *Journal of Parapsychology, 70*, 49–67.

Stanford, R. G. (1974). An experimentally testable model for spontaneous psi events I: Extrasensory events. *Journal of the American Society for Psychical Research, 68*, 34–57.

Stanford, R. G. (1977). Conceptual frameworks of contemporary psi research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 823–858). London, England: Van Nostrand Reinhold.

Stanford, R. G. (1982). An experimentally testable model for spontaneous extrasensory events. In I. Grattan-Guiness (Ed.), *Psychical research: A guide to its history, principles and practices* (pp. 195–205). Wellingborough, UK: Aquarian Press.

Stanford, R. G. (1990). An experimentally testable model for spontaneous psi events: A review of related evidence and concepts from parapsychology and other sciences. In S. Krippner (Ed.), *Advances in parapsychological research 6* (pp. 54–161). Jefferson, NC: McFarland.

Stanford. R. G., & Thompson, G. (1973). Unconscious psi-mediated instrumental response and its relation to conscious ESP performance. *Proceedings of Presented Papers: The Parapsychological Association 16th Annual Convention*, 99-103.

Steinkamp, F. (2005). Forced-choice ESP experiments: Their past and their future. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the 21st century: Essays on the future of psychical research* (pp. 124-163). Jefferson, NC: McFarland.

Steinkamp, E., Milton, J., & Morris, R. L. (1998). A meta-analysis of forced choice experiments comparing clairvoyance and precognition. *Journal of Parapsychology, 62*, 193–218.

Tart, C. (1977). Toward conscious control of psi through immediate feedback training: Some considerations of internal processes. *Journal of the American Society for Psychical Research, 71*, 375–407.

Tart, C. (2001). *Psi: Scientific studies of the psychic realm*. Bloomington, IN: iUniverse.com.

Tart, C. T., Palmer, J., & Redington, D. J. (1979). Effects of immediate feedback on ESP performance: A second study. *Journal of the American Society for Psychical Research, 73*, 151–165.

Acknowledgements

This research was funded by the Bial Foundation grant 105/08. We would like to gratefully acknowledge this support. We should also like to thank the Leslie Church and T. D. Lewis trustees for supporting this project.

Academic Department
UCFB Wembley, Wembley Stadium
Wembley, Middlesex, HA9 0WS, UK
G.Hitchman@ucfb.com