Journal of Scientific Exploration, Vol. 27, No. 3, pp. 393–413, 2013 0892-3310/13

RESEARCH ARTICLE

The Sheep–Goat Effect as a Matter of Compliance vs. Noncompliance: The Effect of Reactance in a Forced-Choice Ball Selection Test

Lance Storm
Brain and Cognition Centre, School of Psychology, University of Adelaide, South Australia 5005, Australia
lance.storm@adelaide.edu.au

Suitbert Ertel

Adam J. Rock

Submitted 8/31/2010, Accepted 3/7/2013, Pre-published 7/15/2013

Abstract—According to Reactance Theory (Brehm & Brehm 1981), when an individual's freedom is threatened through some form of coercion, *reactance* usually sets in. Reactance is "a motivational state aimed at restoring the threatened freedom" (Silvia 2005:277), which may explain the tendency for believers ('sheep') to psi-hit and non-believers ('goats') to psi-miss. In this study, the effect of reactance on psi performance was investigated using Ertel's (2005a, 2005b) Ball Selection Test. It was hypothesized that goats are more reactant than sheep in psi tests because goats are predisposed to *disproving* the psi hypothesis which requires noncompliance. In a laboratory setting, participants completed up to four runs (60 trials/run) of paranormal target-seeking (trying to predict the numbers on table tennis balls). Hit rate for the whole sample ($N = 82$) was significant, 21.06% ($p = .002$), where $P_{MCE} = 20\%$. Participants were randomly assigned to a control condition ($n = 42$) or treatment condition ($n = 40$) requiring them to read a statement that induced reactance. A significant reactance effect was found. There was no significant sheep–goat effect, but the relationship between psi-hit rates and sheep–goat scores was significant. Reactant goats scored significantly lower than control sheep, as expected, but not significantly lower than control goats. Pre-test scores on Tension and Confusion, as measured on The Profile of Mood States—Short Form (POMS-SF) (McNair, Lorr, & Droppleman 1971), predicted psi outcomes.

Keywords: Ball Selection Test—ESP—PK—reactance—psi—sheep–goat effect

Introduction

The Sheep–Goat Effect

Schmeidler (1943, 1945) categorized participants in paranormal experiments as either those who think that ESP is possible under a given experimental condition ('sheep'), or those who reject this possibility ('goats'). The

definition has been extended to include sheep as those who "believe that ESP exists as a genuine phenomenon" (Thalbourne 2003:114), thus excluding goats from this belief. Based on pivotal studies (Lawrence 1993, Palmer 1971, 1977, Schmeidler & McConnell 1973), paranormal belief as measured on sheep–goat scales tends to be a predictor of psi outcomes, with sheep producing significant hit rates (i.e. psi-hitting) and goats producing significant miss rates (i.e. psi-missing). For example, Lawrence's (1993) meta-analysis of 73 studies (totalling 685,000 trials by 4,500 participants) dating back to 1947 demonstrated an accumulative sheep–goat effect that was moderate in size and highly significant—sheep consistently scored better than goats.

As consistent as the sheep–goat effect may be, psi-hitting is assumed to be the product of attempts on the part of sheep to 'prove' the psi hypothesis, but it is also assumed that psi-missing results from goats attempting to 'disprove' the same hypothesis (Palmer 1971, 1972, Schmeidler & McConnell 1973). This principle is captured in Palmer's (1972) 'vindication' theory. Thus, the understanding is that goats have only one agenda (i.e. target avoidance), whereas sheep aim to identify the target. We concur with this premise. While the sheep–goat literature generally indicates that sheep can be successfully encouraged to psi-hit, this evidence comes as a corollary of the fact that sheep have complied with the experimenter's instructions to seek a psi target. Most psi experiments, therefore, are designed to get 'good' performances out of sheep, but they are not designed to get 'bad' performances out of goats. In other words, goats are usually treated like sheep and are instructed to psi-hit; not to psi-miss. Ironically, experimenters then expect a majority of goats to avoid the target, and they even hypothesize psi-missing in goats. What we discovered is that a review of forced-choice studies by Schmeidler and McConnell (1973) showed consistently larger deviations from mean chance expectation (MCE) in sheep compared with goats, with sheep frequently psi-hitting, but goats not as often psi-missing. Although Palmer (1977) reported that 13 out of 17 sheep–goat experiments (76%) from 1947 to 1970 "were in the predicted direction" (p. 193), he did not make any reference to the asymmetrical scoring differences between sheep and goats; a point Steinkamp (2005) later made about Lawrence's (1993) meta-analysis in her review of forced-choice studies. She stressed that Lawrence did not tell us "whether the [sheep–goat] difference is due, for example, to goats tending to perform significantly badly, with sheep scoring at chance, or to sheep performing significantly well with goats scoring at chance (or something in between these two alternatives)" (pp. 152–153).

To validly test the counter-position of goats, experimenters should not

simply instruct goats to psi-hit and then measure the shortfall in hits—they must either encourage goats to psi-miss, or *discourage* them from psi-hitting. However, these attempts are likely to fail, which is probably why no experimenter does it, because goats do not seek to 'prove' the psi hypothesis in any of its forms, as psi-hitting *or* psi-missing. Clearly, seeking any kind of compliance from goats will not succeed.[1] When asked to comply in any manner, goats will probably perceive the request as a threat to their freedom, which may cause so-called 'boomerang effects' (i.e. noncompliance and its consequences). For that reason, we regard the motivation of goats as being a special case of reactance behavior (a kind of resistance or refusal to cooperate—see next section) distinct from the kind of motivation that impels sheep to comply. While Schmeidler and McConnell (1973) intimated a sheep–goat distinction, they only ever 'supposed' the causes, nor did they adequately explain the asymmetrical deviations around the MCE of sheep and goats (the very same point to which Steinkamp drew our attention, as mentioned above). If psi-missing in goats is the effect parapsychologists are looking for, we argue that experimenters must manipulate goats' behavior and thus produce psi-missing, and one way to do this is to induce reactance in goats. If this manipulation is successful, then the assumption is demonstrated that goats have a motive that is antithetical to that of sheep. These ideas are developed further in the next section.

Reactance Theory and Goats' Behavior

In the field of social psychology, it is found that if a person's attitudinal or behavioral freedom is threatened or reduced, the person becomes motivationally aroused (Kraus 1995, Smith 1978, Worchel & Brehm 1970, Wright 1986). This arousal generates psychological *reactance* that even takes the form of 'boomerang effects' where the individual adopts a noncompliant attitude, or engages in noncompliant behavior on the assumption that freedom will be restored. Considerable work has been done testing *reactance theory* (Brehm 1966, Brehm & Brehm 1981, Dillard & Shen 2005, Miller, Burgoon, Grandpre, & Alvaro 2006, Miller, Lane, Deatrick, Young, & Potts 2007, Silvia 2005, 2006). For example, Silvia (2005) showed that reactance was highest in the group that felt the most threatened by the content of an opinionated communication (i.e. a reactance prime). Opinionated communications are the most widely used threat manipulations in reactance research (Brehm & Brehm 1981). In a follow-up study, Silvia (2006) found that disagreement directly motivated by a threat declined when the threat was removed.

To test reactance theory on goats, parapsychologists would also need a reactance treatment (i.e. reactance prime) in the form of an opinionated

communication. According to theory, the treatment will raise reactance, which will remain high if no outlet is provided, and since there is a relationship between attitude and behavior (Ajzen 1985, Kraus 1995), we can expect increased noncompliant behavior in goats under threat, which may thus yield increased target *avoidance* and therefore shifts from chance-scoring to psi-missing.

While there have been no parapsychological studies on reactance *per se*, Lovitts (1981) did investigate the sheep–goat effect in a related way. He divided participants into two groups—one comprising participants who were told they were in an experiment to demonstrate ESP ability, and another which was told subliminal perception was a legitimate (non-paranormal) theory of ESP. A significant interaction effect was found indicating that the effects of the two conditions on psi-scoring were not the same across levels of paranormal belief (i.e. sheep appeared to have been manipulated to score like goats, and vice versa). The problem with Lovitt's design was that the second (subliminal perception) group did not know they were in a psi experiment. Lawrence (1990–1991) also designed an experiment that was an explicit attempt to manipulate psi outcomes by expressly telling participants that the test was designed to prove ESP, or disprove ESP, depending on random assignment. Lawrence failed to replicate the effect, but at least both groups knew they were in a psi experiment.

Evidence that goats can be manipulated into changing their psi performance comes from a study by Storm and Thalbourne (2005). Storm and Thalbourne's objective was to see if skepticism in some number of goats could be manipulated to such a degree that they would 'convert' from a psi-dismissive disposition to a psi-supportive disposition. It was hypothesized that naïve goats, after having the implications of significance testing explained to them, would thereby adopt a newfound belief in psi, so that psi outcomes would shift from chance-scoring (or psi-missing) before the manipulation to psi-hitting after the manipulation. The hypothesis was supported—in a symbol identification task, goats shifted from chance-scoring (20%, where P_{MCE} = 20%), to psi-hitting (30%, p = .047). The significant change in performance was referred to as a 'conversion effect'.

So far, however, there have been no studies that directly test the manipulation of reactance in goats. In alignment with the conventional view in parapsychology, we advocate the position that goats show an aversion to psi-hitting. In order that reactance effects can be tested in sheep and goats, the Australian Sheep–Goat Scale (ASGS) (Thalbourne 1995) was used in the present study to differentiate sheep from goats. The psi test used was Ertel's (2005a) Ball Selection Test. We chose this test because it offers a degree of ecological validity not found in most laboratory psi tests, and the

highly physical nature of its procedure may be particularly motivating of reactance effects not only in attitude but also in behavior. The Ball Test is now described.

The Ball Selection Test

In Ertel's (2005a, 2005b, 2010) standard Ball Selection Test, participants are handed an opaque bag containing 50 table tennis balls on which one of the numbers 1 to 5 are written, each number on 10 balls. A participant's trial consists of shaking the bag, blindly selecting (i.e. drawing out) a ball from the bag by hand through the small opening of the bag, checking the number, and putting the ball back in the bag. They are told to look away from the bag when they select a ball, and they are under observation by the experimenter during all trials to make sure they are not looking into the bag to see the balls. On each trial, the number is guessed in advance and recorded. In the same trial, the number on the ball that is selected is also recorded.

Studies by Ertel (2005a, 2005b) have produced significant results using the Ball Test. In a sample of 231 students (Ertel 2005b), the average hit-rate deviation was 9% above the MCE, which is extremely significant, $z = 12.07$ ($p < 10^{-15}$). In a follow-up study (Ertel 2005a), a number of tests were run. One test unit consisted of six or eight runs comprising 60 trials each (total: 360 or 480 trials). Sixteen high-scoring participants of Sample 1 were also tested under laboratory control, again using the Ball Test procedure (Version I). Sample 2 took the Ball Test (Version II) which is essentially the same as Ball Test Version I except that green or red dots are sprinkled over the balls, and participants guess numbers (five targets) *and* colors (two targets), where P_{MCE} is thus 10%. Thirteen high-scorers of Sample 2 were also tested under laboratory control using the bead-selection test where each participant selected one of five colors (no numbers, $P_{MCE} = 20\%$). As hypothesized, hit-rates of high scorers under laboratory control were significantly above chance.

In the latest study (Ertel 2010), using the Version II design, nine selected participants retested in a laboratory achieved a highly significant hit rate of 17.3% ($P_{MCE} = 10\%$). Also, a replication of the laboratory procedure was conducted by two graduating students working under the guidance of a skeptical professor at the Anomalistic Psychology Research Unit (APRU) at Goldsmiths College, London University. Their 40 unselected APRU participants achieved a hit rate of 10.8% (where $P_{MCE} = 10\%$), which was significant ($p = .002$). It is important to note that Ertel (2005b, 2010, 2013) has gone to considerable lengths to demonstrate that psi-hitting using the Ball Test design is not an artifact of sensory leakage, memory leakage, or randomization failure (see especially Ertel 2013).

Mood States, Enthusiasm, and Psi

There has been little research on the effect of mood states on paranormal performance (e.g., Carpenter 1991, Crandall 1985, Schmeidler 1988). This oversight may stem from the fact that most psi researchers have focused on the broader issues of psi-modifying effects of attitude and personality (for an early review, see Palmer 1977), or the psychopathological aspects of paranormal belief and experience (e.g., Thalbourne & Delin 1994; see also Irwin 2009 for a review). However, it follows that a participant's enthusiasm and motivation for a psi outcome is bound to be reflected in his/her mood, and parapsychologists have acknowledged this relationship (Carpenter 1991, Stanford 1977).

Some researchers (Stanford 1977) note that motivation is a broad concept, difficult to specify, since it can refer to psychophysiological factors such as "emotional arousal," or cognitive factors such as "attention-focusing" (p. 842). At least one parapsychologist (Carpenter 1991) saw no difficulty in measuring mood, though he worked on a succession of scales, each one being an improvement on its predecessor. Ultimately, he found that psi-hitting was associated with moods whereby the participant felt strong-willed, detached, and agreeable, whereas psi-missing was associated with anxiety. He also found that large run-score variance was associated with carefree moods, whereas small variance was related to annoyance.

Since 1971, the Profile of Mood States (POMS) assessment has been shown to be an excellent measure of the mood states Tension, Depression, Anger, Vigor, Fatigue, and Confusion, plus an aggregate scale, Total Mood Disturbance, which is a combined score of five scales (Tension, Depression, Anger, Fatigue, and Confusion) minus the score for Vigor (McNair, Lorr, & Droppleman 1971). Although the POMS is applied mostly in clinical and therapeutic situations, it is suitable in research where experimental manipulations are required and mood changes might need to be assessed in order to understand the nature of the effects under investigation. We did not find any parapsychological studies that employed the POMS. However, since we are interested in the influence of the psychological state of reactance on psi performance and its related psychological effects, we planned to extend our inquiry by including measures of the six moods listed above, as well as enthusiasm, to determine any differences between the reactance treatment group and the control group on these variables. We expected negative shifts in mood due to the reactance treatment.

We were also interested in whether enthusiasm scores and the six mood scale scores predict psi performance, and whether there are sheep–goat differences in mood. In addition, we assessed participants' opinions and perceptions of the Ball Test by asking them two questions: "Do you think

your result will prove that psi exists?", and "Assuming psi exists, is this test suitable for psi assessment?" These questions are different ways of looking at paranormal belief. The ASGS (mentioned above) is a broader measure that takes into account a number of factors about psi beliefs and/or psi experiences, whereas the two ('prove psi' and 'suitability') questions pertain to the experimental design. Our two questions actually reflect Schmeidler's (1943, 1945) original definition, so it was of interest to determine if scores on the ASGS would be related to participants' attitudes toward the experimental design.

Hypotheses

1. Main Effect: There is a psi-scoring difference between the reactance treatment group ('reactants') and the control group ('controls'), where the reactance group scores lower than the control group.
2. Main Effect: There is a sheep–goat effect on psi-scoring, where sheep score higher than goats.
3. Interaction Effect: There is an interaction effect, where the effect on psi of the reactance treatment is not the same across levels of belief (i.e. between sheep and goats). Specifically, reactant goats have the lowest psi scores, and control sheep have the highest psi scores.
4. There is a positive relationship between psi-hit rates and Rasch-scaled Australian Sheep–Goat Scale (RASGS) scores.
5. There are mood changes (i.e. differences between pre-POMS and post-POMS scores) in POMS Tension, Depression, Anger, Fatigue, Confusion, and the aggregate scale Total Mood Disturbance, due to the reactance treatment, with greater reductions for controls compared with reactants (Vigor was excluded in the analysis).
6. Pre-test scores on POMS Tension, Depression, Anger, Vigor, Fatigue, Confusion, and Total Mood Disturbance correlate with psi-hit rates.
7. Pre-test scores on POMS Tension, Depression, Anger, Vigor, Fatigue, Confusion, and Total Mood Disturbance are different between sheep and goats.
8. Level of Enthusiasm is lower for reactants compared with controls.
9. Psi-hit rates correlate positively with (i) answers to the Psi Proof question; (ii) answers to the Ball Test Suitability question; and (iii) Enthusiasm.
10. RASGS scores correlate positively with (i) answers to the Psi Proof question; (ii) answers to the Ball Test Suitability question; and (iii) Enthusiasm.

Tests used include the Binomial test, ANOVA, *t* test, and Pearson's *r*.

Method

Participants

The sample mainly comprised students from Deakin University, Melbourne, Victoria. Students did not participate as part of their course requirement, or to gain course credit. Participants were 'unselected' with the aim of recruiting an even number of sheep and goats. A total of 82 participants were laboratory-tested. Mean age for a reduced sample ($N = 79$), since three participants did not give their age, was 29 years ($SD = 13$ years). Fifty-four percent of the sample were females ($n = 44$; males: $n = 38$, 46%).

Materials

Eight forms and measures were used in the study: (1) *Instruction Sheet*; (2) *Consent Form*; (3) *Score Sheet* (comprising four grids, 15 × 4 cells, = 60 trials); (4) Thalbourne's (1995) 18-item Rasch-scaled *Australian Sheep–Goat Scale* (RASGS) (Lange & Thalbourne 2002). An 18-item scale measures belief and alleged experience of paranormal phenomena. Each item scores 0, 1, or 2 points, where 0 = *False*, 1 = *Uncertain*, and 2 = *True* (Raw range is 0 to 36; Raw $M = 18$). The ASGS data are then top-down purified (two items are removed; items #9 and #10) to eliminate age and gender bias from the scale (Lange & Thalbourne 2002), and this procedure alters the scoring range and mean; (5) The Profile of Mood States—Short Form (POMS-SF) (McNair, Lorr, & Droppleman 1971) is a 30-item self-report measure of positive and negative mood states. Each item represents an affective state (e.g., Shaky, Grouchy, Annoyed, Lonely, Weary, etc.), and respondents are required to indicate how much each item represents their present state on a scale from 0 = *Not at all* to 4 = *Extremely*; (6) The single-item rating-scale on psi proof: "Do you think your result will prove that psi exists?" Participants respond on a 5-point Likert scale ranging from 0 = *Definitely-No* to 4 = *Definitely-Yes*; (7) The single-item rating-scale on the Ball Test Suitability: "Assuming psi exists, is this test suitable for psi assessment?" Participants respond on a 5-point Likert scale ranging from 0 = *Strongly Disagree* to 4 = *Strongly Agree*; and (8) The single-item Enthusiasm (before every run): "On a scale of 1 ('low') to 5 ('high'), my enthusiasm is ___".

Apparatus

An opaque bag (30 cm × 40 cm) containing 50 table tennis balls was used. One of five digits (1, 2, 3, 4, or 5) is written on each ball, evenly spread on its surface. There are 10 balls bearing each digit (Total: 50 balls).

Procedure

All participants (N = 82) read the Instruction Sheet, and they signed the Consent Form if they decided to continue with the experiment. They completed the ASGS, and the POMS (for pre-test scores), and then read the opinionated communication—a modified version of the one used by Silvia (2005). The communication was presented surreptitiously as basic information about participating in a psi study:

> This short communication was written by a university professor: This Ball Test has been developed in Germany and I claim that it is the best of all procedures that have hitherto been applied in parapsychology. I am utterly convinced that psi exists and that participants cannot avoid letting their psi power come to the fore when they make number predictions and then select the numbered balls. Every person, I claim, is expected to display such power. *I know I have persuaded you about this. I know you agree with my opinion. In fact, you're really forced to agree because university students can't have differing opinions on this issue.*

The additional italicized sentences were intended to function as the threatening elements; they were not italicized in the study. In the control condition, participants received the same communication without the italicized sentences. Allocation to control or treatment group was random. The two types of questionnaire, control and treatment, were originally stacked in two piles and then, using a random number table, randomly stacked one by one into a single pile. The participant was automatically and randomly assigned to a group according to the questionnaire that was drawn from the top of the pile. The questionnaire cover page did not identify whether it was for a 'control' or 'treatment' participant.

Participants rated their enthusiasm just prior to starting the Ball Test, and then they were shown how to shake the bag, and how to select the balls (on behalf of participants, the experimenter [LS] recorded the outcomes directly onto the Score Sheet). Enthusiasm was rated prior to every run so up to four enthusiasm ratings were taken for participants who completed four runs.

Each participant completed up to four runs of 60 trials in one laboratory session, though many did not complete four runs. According to Deakin University's Human Research Ethics Committee policy, participants must be given the option to withdraw at any stage. Accordingly, the number of completed runs varied among participants. Some participants did not have time for four runs; others got bored. Most reasons for withdrawal were of that nature. It is well understood in parapsychology that it is better to let participants withdraw if they feel bored, disinterested, or unmotivated, as

these attitudes allegedly undermine psi processes and can result in chance-scoring or psi-missing (see Broughton 1991, Schmeidler 1988, Stanford 1986). These withdrawals could not be considered feedback-based optional stopping because feedback was given to participants only after their last run was completed by which time they had already made it clear that they did not want to start another run.

The total testing time for 4 × 60 trials = 240 trials, spread over one session, amounted to approximately one hour. The MCE for a run of 60 trials is 12 hits because with five numbers of which participants guess one number, 12 hits (i.e., 60/5) will be drawn on average. For 240 trials, the MCE = 48 hits (i.e. 360/5).

At the end of psi testing, participants completed the POMS again (for post-test scores), and they answered the two questions: "Do you think your result will prove that psi exists?" and "Assuming psi exists, is this test suitable for psi assessment?"

Participants were told that when the test was done and the data analyzed, they would receive feedback in the form of a general information flyer distributed in the Deakin University library (no confidential information was provided).

Results

Descriptive Statistics

The Ball Test. There were a total of 12,016 trials of which there were 2,531 hits (21.06%). A total of 82 participants completed 224 runs of 60 trials each. Of the 81 participants who completed at least one run, 28 participants (35%) completed 4 runs; 11 participants (13%) completed 3 runs; 13 participants (16%) completed 2 runs, and 29 participants (36%) completed 1 run. One participant did not complete the first run (only 16 trials were attempted).

As a preliminary test on participation rates for the control (n = 42) and reactance (n = 40) groups, we found that the fewest number of runs were completed by the reactants (91 runs), whereas the control group completed 109 runs. However, the difference was not significant, $\chi^2(1, N = 82) = 4.86$, $p = .183$ (two-tailed).

Rasch-Scaled Australian Sheep–Goat Scale (RASGS). The mean score for the raw-score version of the ASGS was 15.00 (SD = 8.21). The skew of the distribution of scores was normal (skew = –0.22, SE = 0.27). The mean score for the Rasch-scaled version (i.e. RASGS) was 24.46 (SD = 6.64). The theoretical range and observed range were the same, 8.13 to 43.39. Reliability of the ASGS was high: Cronbach's alpha = 0.93 (Cronbach's

alpha was also 0.93 for the RASGS). Neither of the two demographic variables, Age and Sex, correlated with RASGS. The median score (26.10) was used as the cut-off point between sheep and goats, with goats scoring 26.10 or lower (n = 42), and sheep scoring above 26.10 (n = 41). Goats' mean RASGS score = 19.72 (SD = 5.23) and sheep mean RASGS score = 29.57 (SD = 3.44). The difference was significant, $t(71.43) = -10.08$, $p < .001$.

Profile of Mood States (POMS). The POMS measures seven variables: Tension, Depression, Anger, Vigor, Fatigue, and Confusion, plus a composite measure, Total Mood Disturbance. Pre-test and post-test scores are presented in Table 1. Note that scores on all seven measures decreased pre- to post-testing, but the causes for these decreases are considered below (see Hypothesis 5). For the whole sample, there were significant differences between pre-test and post-test scores on (i) Tension, $t(65) = 3.08$, $p < .003$ (two-tailed); (ii) Depression, $t(65) = 3.04$, $p < .003$ (two-tailed); and (iii) Confusion, $t(79) = 2.28$, $p < .026$.

TABLE 1
Mean Scores and *SD*s for Profile of Mood States Seven Variables (*N* = 66)

POMS Variables	Pre-Test		Post-Test	
	Mean	*SD*	Mean	*SD*
Tension	3.32	3.22	2.36	3.72
Depression	2.71	3.62	1.92	3.20
Anger	1.39	2.19	1.08	2.06
Vigor	7.37	4.13	7.04	4.38
Fatigue	5.55	4.82	5.29	4.74
Confusion	3.87	2.56	3.37	2.71
Total Mood Disturbance	9.46	13.05	6.98	12.04

Enthusiasm. Scores ranged between 0 and 5, which is also the theoretical range, mean score = 3.47 (SD = 0.89); median score = 3.50. Neither Age nor Sex correlated with Enthusiasm.

Psi Proof Question. Scores ranged between 0 and 4, which is also the theoretical range; mean score = 1.73 (SD = 1.04); median score = 2.00. Neither Age nor Sex correlated with the question.

Ball Test Suitability Question. Scores ranged between 0 and 4, which is also the theoretical range; mean score = 2.53 (*SD* = 1.04); median score = 3.00. Neither Age nor Sex correlated with the question.

Planned Analyses

H1: There is a psi-scoring difference between the reactance group and the control group, with the reactance group scoring lower than the control group. Table 2 shows numbers of hits and trials, and hit rates as proportions for the whole sample ($N = 82$), the reactance group ($n = 40$), and the control group ($n = 42$). The overall hit rate for the whole sample was significant, 21.06% (binomial $z = 2.90$, $p = .002$; $ES = .03$). There was a significant reactance effect on proportion of hits, with 'reactants' (mean percentage = 20.26%) scoring significantly lower than 'controls' (mean percentage = 21.74%), $F(1, 77) = 2.75$, $p = .05$ (one-tailed). Table 2 also shows that the controls produced a significant hit rate ($p < .001$), but reactants did not ($p = .323$). The hypothesis was supported.

It may be hypothesized that decline effects might have occurred, with low scores predominating in the reactance condition relative to the control condition, so we also conducted analyses for first-run data only (tests on first-run data were also conducted for *H2* and *H3*). There was a significant reactance effect on proportion of hits, with reactants (mean percentage = 19.92%, $p = .468$) scoring significantly lower than 'controls' (mean percentage = 22.03% $p = .006$), $F(1, 77) = 2.93$, $p = .04$ (one-tailed). The hypothesis was again supported.

We also looked at participant-based outcomes in the form of *z* scores and *ES* values. The mean *z* score for the whole sample was very weak at $z = .0002$ ($ES = 0.03$). The sheep *z* score was 0.08 ($ES = 0.01$), and goats' *z* score was 0.039 ($ES = 0.03$).

H2: There is a sheep–goat effect on psi-scoring, with sheep scoring higher than goats. Table 2 also shows that sheep and goats both produced significant hit rates. The sheep hit rate (21.26%) *was* higher than the hit rate for goats (20.89%). Though hit rates were in the direction hypothesized, with the sheep hit rate higher than goats, there was no significant sheep–goat effect, $F(1, 77) = 1.07$, $p = .152$ (one-tailed). The hypothesis was not supported. This nonsignificant sheep–goat effect may be due to low power as four groups had to be formed for the univariate ANOVA test based on two dichotomous variables: (i) reactance/control and (ii) sheep/goat (see the contrasting result for Hypothesis 4 below).

Once again, when we restricted the analysis to only the first-run data, the sheep hit rate (21.45%, $p = .042$) *was* higher than the hit rate for goats

TABLE 2
Number of Trials, Hits, and Hit Proportions: Reactants and Controls, and Sheep and Goats

Group	Trials	Hits	Hit Proportion	*p* (one-tailed)
Reactance (*n* = 40)	5460	1106	20.26%	.323
Control (*n* = 42)	6556	1425	21.74%	< .001
Total (*N* = 82)	12016	2531	21.06%	.002
Goats (*n* = 42)	6060	1266	20.89%	.042
Sheep (*n* = 39)	5940	1263	21.26%	.008
Total (*N* = 81)*	12000	2529	21.08%	.002

* One participant did not complete the ASGS.

(20.55%, $p = .251$). Though hit rates were in the direction hypothesized, with the sheep hit rate *higher* than goats, there was no significant sheep–goat effect, $F(1, 77) = 0.71$, $p = .201$ (one-tailed). The hypothesis was not supported.

H3: There is an interaction effect, where the effect on psi of the reactance treatment is not the same across levels of belief (i.e. between sheep and goats). Specifically, reactant goats have the lowest psi scores, and control sheep have the highest psi scores. There was no significant interaction effect, $F(1, 77) = 0.003$, $p = .959$ (one-tailed). However, at 19.42%, reactant goats ($n = 19$) did score the lowest of all four sub-groups. Control goats scored at 21.01% ($n = 23$), reactant sheep scored at 20.40% ($n = 21$), and control sheep scored at 22.09% ($n = 18$), which was expected according to reactance theory applied to goats. Mean hit rate for the reactant goats, however, was not significantly lower than for any other group, $F(3, 80) = 1.18$, $p = .161$ (one-tailed). Also, control sheep scored the highest. The difference between reactant goats and control sheep was significant, $t(35) = –1.76$, $p = .034$ (one-tailed). The hypothesis was partially supported.

Again, when we restricted the analysis to first-run data, there was no significant interaction effect, $F(1, 77) = 0.003$, $p = .477$ (one-tailed). However, at 19.39%, reactant goats ($n = 19$) did score the lowest of all four sub-groups (control goats scored at 21.52% [$n = 23$]; reactant sheep scored at 20.40% [$n = 21$]; and control sheep scored at 22.69% [$n = 18$]), which was expected according to reactance theory applied to goats. The mean hit rate for reactant goats, however, was not significantly lower than for any other

group, $F(3, 80) = 1.14$, $p = .170$ (one-tailed). Also, control sheep scored the highest again. The difference between reactant goats and control sheep was again significant, $t(35) = –1.80$, $p = .040$ (one-tailed). The hypothesis was partially supported. There tests on first-run data only in Hypotheses 1, 2, and 3 indicate that decline effects do not explain the differences between the two groups, reactance and control.

H4: There is a positive relationship between psi-hit rates and RASGS scores. The relationship was positive and significant, $r(79) = 0.20$, $p = .036$ (one-tailed). The hypothesis was supported.

H5: There will be mood changes in POMS Tension, Depression, Anger, Fatigue, Confusion, and Total Mood Disturbance, due to the reactance treatment, with greater reductions for controls compared to reactants. Difference scores between POMS 'pre' and 'post' psi testing were calculated for reactants and controls (Vigor was excluded). To minimize error variance, we used a mixed-model ANOVA with the sheep–goat group added as a between-subjects variable. There was only one effectively significant difference: Anger, $F(1, 62) = 2.64$, $p = .054$ (one-tailed), but given that this test is only one of six tests that were run, the outcome may be a chance effect. The hypothesis was not strongly supported.

H6: Pre-test scores on POMS Tension, Depression, Anger, Vigor, Fatigue, Confusion, and Total Mood Disturbance correlate with psi-hit rates. Psi-hit rates correlated positively and significantly with Tension, $r(80) = .29$, $p = .009$ (two-tailed); and Confusion, $r(80) = .33$, $p = .002$ (two-tailed). Two tests out of seven (29%) is well above the MCE (following the 5% rule), since we would only expect about one test in seven to be significant by chance alone (note that applying a Bonferroni correction, given that there were seven tests, still yields p values less than .05). Although the hypothesis was partially supported, it is not clear why these two negative moods would facilitate the psi function (see the Discussion for comments on this issue).

H7: Pre-test scores on POMS Tension, Depression, Anger, Vigor, Fatigue, Confusion, and Total Mood Disturbance are different between sheep and goats. There were no significant differences between sheep and goats. The hypothesis was not supported.

H8: Level of Enthusiasm is lower for reactants compared to controls. Although the reactance group did have a lower mean Enthusiasm score ($M = 0.98$) compared with the control group ($M = 1.02$), Enthusiasm in

TABLE 3
Enthusiasm Ratings (First Run): Reactants vs. Controls

Group	Enthusiasm Score for First Run					Total	Mean Score
	1	2	3	4	5		
Reactance (n = 40)	1	3	16	14	5	39	0.98
Control (n = 42)	1	1	18	18	4	42	1.02

the reactance group was not significantly lower than in the control group, $t(80) = 0.69$, $p = .248$ (see Table 3). The hypothesis was not supported.

H9: Psi-hit rates correlate positively with (i) answers to the Psi Proof question; (ii) answers to the Ball Test Suitability question; and (iii) Enthusiasm. The Psi Proof and Ball Test Suitability questions were asked at the end of psi testing. Enthusiasm was measured before each of up to four runs, though not all participants completed four runs. The number of participants rating Enthusiasm prior to Run 1 was much larger than on the other three runs, so only Enthusiasm data for the first run was used. Results of the three-part hypothesis are as follows: (i) The Psi Proof question: The relationship was positive, but weak and not significant, $r(80) = 0.03$, $p = .382$ (one-tailed); (ii) The Suitability question: The relationship was positive and significant, $r(80) = 0.45$, $p < .001$ (one-tailed); (iii) Enthusiasm: The relationship was positive but only approached significance, $r(80) = 0.17$, $p = .063$ (one-tailed). The hypothesis was partially supported.

H10: RASGS scores correlate positively with (i) answers to the Psi Proof question; (ii) answers to the Ball Test Suitability question; and (iii) Enthusiasm. Results of the three-part hypothesis are as follows: (i) The Psi Proof question: The relationship was positive and significant, $r(78) = 0.38$, $p < .001$ (one-tailed); (ii) The Suitability question: The relationship was not positive, nor was it significant, $r(78) = –0.12$, $p = .137$ (one-tailed); (iii) Enthusiasm: The relationship was positive and significant, $r(79) = 0.26$, $p = .009$ (one-tailed). The hypothesis was partially supported.

Discussion

Using the Ball Selection Test as a means by which psi effects might be elicited, the present study sought to gain insight into the psi performance differences between sheep and goats after a reactance manipulation. The

conventional understanding is that sheep are compliant toward psi-hitting, whereas goats are noncompliant and try to avoid targets, and thus may psi-miss. This conventional design means sheep are encouraged to do their 'best', but goats are not encouraged to do their 'worst'. Following the principles of reactance theory (Brehm & Brehm 1981), we deemed it possible that psi activity in goats could be changed by manipulating their reactance with an opinionated communication, resulting in even worse psi performances than would be expected of them under normal (control) conditions.

We showed that reactance does have an effect on psi performance, with the treatment clearly showing adverse effects on psi-hitting (Hypothesis 1). The mainly non-significant results for Hypothesis 5 suggest that reactance was the causal factor underlying the psi performance change, given that the consensus in the social sciences is that the treatment is the *cause* of performance differences, all things being equal. We therefore claim that the reactance treatment caused a change in psi performance. Our study is the first parapsychological study to contribute to the literature which finds that threatening communication treatments affect reactance, which then affects task performance (Brehm & Brehm 1981, Silvia 2005, 2006, Smith 1978).

Another of our primary aims was to show that the treatment would have more of an effect on psi-scoring by goats than psi-scoring by sheep (Hypothesis 2). This effect was not shown directly, although scoring was in the direction hypothesized (we attribute this non-significant difference to low power in the univariate ANOVA test, since we did find a significant sheep–goat effect when we tested Hypothesis 4). With no other explanation for the failure, and assuming our hypothesis is true, we surmise that 'trait' reactance may be relatively high in goats at the pre-experimental stage (i.e. before testing began, which is before reading the communication) compared with sheep (i.e. there may be a ceiling effect on reactance for goats). If there is less latitude in goats to increase reactance compared with sheep, it makes good sense to control for trait reactance. In future psi tests on reactance, it is advised that the covariate of trait reactance at the pre-experimental stage be measured so that it can be controlled in statistical tests. To that end, a replication study is planned that will feature the Hong Psychological Reactance Scale (Hong & Faedda 1996).

Note, however, that we twice found a significant scoring difference between control sheep and reactant goats (Hypothesis 3), with reactant goats scoring lower than control sheep. Also, psi-scoring for reactant sheep and control sheep was not significantly different, and scoring for reactant goats and control goats was also not significantly different. The only significant difference was between reactant goats and control sheep, which is a sheep–goat effect modified by reactance.

Although we did not find a significant difference in psi-hit rates between sheep and goats (Hypothesis 2), we did find a significant positive relationship between RASGS scores and hit rates (Hypothesis 4). This latter effect is very much expected as it is another way of looking at the sheep–goat effect, and the correlate is quite often tested in psi studies. Thus we report that a sheep–goat effect has been demonstrated using a more sensitive (continuous) measure of the full RASGS scale rather than a discrete sheep–goat dichotomy based on an arbitrary split into two groups (i.e. sheep and goats) based on a median score.

In Hypothesis 5, we tested the psychological effect of reactance on six mood states as measured on the POMS: Tension, Depression, Anger, Fatigue, Confusion, and Total Mood Disturbance (Vigor was excluded). There was some evidence that the reactance treatment hampered reductions in Anger, suggesting that reactants were as stable as controls over the duration of the experiment. At this early stage, it is still feasible that the opinionated communication has effects on factors other than psi-hitting since reactance is a psychological response, and multiple psychological responses are usually elicited even though only one stimulus is presented. This assumption is borne out by the findings of Hypothesis 6, which indicate that psi-scoring is related to levels of tension and confusion. In that instance, psi-hit rates correlated positively and significantly with the variables Tension and Confusion. Two tests out of seven (29%) is well above the MCE (following the 5% rule), since we would only expect about one test in seven to be significant by chance alone (note that applying a Bonferroni correction, given that there were seven tests, still yields p values of less than .05).

Although Hypothesis 6 was partially supported, it is not clear why increases in these two negative moods, Tension and Confusion, would facilitate the psi function. As we used pre-psi test scores of the mood measures, it may be that participants had doubts, expressed as tension and confusion, as to what was expected of them during the experiment, even though the Instruction Sheet and experimenter's subsequent explanations were intended to make it clear. If it is thought that these results might suggest a sheep–goat effect, we did not find any pre-test differences between sheep and goats on any of the POMS variables (Hypothesis 7). These null findings are useful as they indicate that sheep and goats were essentially matched on these POMS variables at the start of the experiment, thus rendering redundant the claim that findings of tests of Hypotheses 5 and 6 are artifacts of psi belief. However, resolving this issue may be problematized by Carpenter's (1991) finding that psi-missing tends to be associated with anxiety, which may be similar to tension—note that we found a significant correlation between POMS Anxiety and Tension

before the psi test, $r(80) = .44$, $p < .001$ (two-tailed), and after the psi test, $r(64) = .53$, $p < .001$ (two-tailed). We note that Tension can be a measure of motivation rather than dysfunctional anxiety.

We found no significant difference between reactants and controls on enthusiasm (Hypothesis 8); the correlation of psi-hitting with enthusiasm approached significance (Hypothesis 9); and enthusiasm correlated with RASGS scores (Hypothesis 10). Recall also that psi-hitting correlated significantly with RASGS scores (Hypothesis 4). Thus, those participants who were more enthusiastic tended to be sheep, and also tended to score better on the psi task, though we only have suggestive evidence that sheep scored better than goats, since we found no differential sheep–goat effect (Hypothesis 2), only a relationship with RASGS scores and psi-hitting (Hypothesis 4).

Psi-hitting did not correlate significantly with 'psi proof', but it did correlate significantly with 'suitability' (Hypothesis 9), so the higher the psi score, the more participants thought the Ball Test was a suitable test of psi. Note, however, when the sheep–goat measure is considered, we find a reversal of effect. RASGS scores correlated significantly with 'psi proof', but did not correlate significantly with 'suitability' (Hypothesis 10). Sheep, as believers, would be more inclined than goats to accept that psi-hitting scores 'prove' psi, but it is ironic that psi scores did not predict 'psi proof' as it is likely that psi-hitters were mostly sheep (perhaps sufficient numbers of sheep with high scores were too modest to say 'Yes' to the 'psi proof' question, and most goats are likely to say 'No' anyway). It is important to note that most participants had prior knowledge of their psi scores having been given feedback after each run, and then they answered the two ('psi proof' and 'suitability') questions. We point out, however, that getting feedback of a total score out of 15 for each and every run is one thing; being told that a score was significantly above the MCE or not above the MCE is another, and we stress that participants were never told whether or not they psi-hit. Nevertheless, the test results for Hypotheses 9 and 10 may be spurious, if not ambiguous, with the only way of validly testing these relationships being to keep participants blind to the scoring component until the end of the experiment when all questions are answered. This protocol, however, is inadvisable. For example, Honorton and Ferrari (1989), in their meta-analysis, showed that the largest psi effects were found in forced-choice studies where feedback was given. We must also allow for the fact that our single-item questions by their nature (i.e., 'psi proof' and 'suitability') may be unreliable and cannot be treated with the same regard as our other two measures, the RASGS and the POMS, both of which have good psychometric properties. The same assumption may apply to the

Enthusiasm question. For that reason, findings and conclusions based on test results on Hypotheses 8, 9, and 10 must be treated with due caution.

Conclusion

The findings in this study show that reactance does influence psi in a mixed sample, but there are differential effects when we consider sheep and goats separately. Although reactant goats performed significantly worse than control sheep, it may not suffice to manipulate state reactance in order to elicit a reactance-related psi effect in goats, because goats may have a higher mean level of pre-experimental 'trait' reactance compared with sheep. It is planned to measure trait reactance in a follow-up study so that pre-experimental reactance can be controlled.

We also found that the reactance treatment may have hampered reductions in Confusion and Depression over the course of the experiment, but we also found that Tension and Confusion predicted hit rates. Finally, we note that the reactance treatment did not affect Enthusiasm—that being said, we did find evidence that the more enthusiastic participants also tended to be sheep, who also tended to score better on the psi task.

Note

[1] Of course, some proportion of goats might try to comply, which still suggests that other goats (probably the majority) are noncompliant, and will then try to avoid the target. In the same sense, the majority of sheep would be compliant.

Acknowledgements

Preparation of this article was supported by a grant from the Bial Foundation. The authors thank Fiona Permezel for her assistance in recruiting some participants.

References Cited

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In *Action Control: From Cognition to Behavior* edited by J. Kuhl and J. Beckman, Heidelberg, Germany: Springer, pp. 11–39.

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. New York: Academic Press.

Brehm, S. S., & Brehm, J. W. (1981). *Psychological Reactance: A Theory of Freedom and Control*. New York: Academic.

Broughton, R. S. (1991). *Parapsychology: The Controversial Science.* New York: Ballantine.

Carpenter, J. C. (1991). Prediction of forced-choice ESP performance. Part III: Three attempts to retrieve coded information using mood reports and a repeated-guessing technique. *Journal of Parapsychology, 55*, 227–280.

Crandall, J. E. (1985). Effects of favourable and unfavourable conditions on the psi-missing displacement effect**.** *Journal of the American Society for Psychical Research*, *79*, 27–38.

Dillard, J., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, *72*, 144–168.

Ertel, S. (2005a). Psi test feats achieved alone at home: Do they disappear under lab control? *Australian Journal of Parapsychology*, *5*, 149–164.

Ertel, S. (2005b). The Ball Drawing Test: Psi from Untrodden Ground. In *Parapsychology in the 21st Century: Essays on the Future of Psychical Research* edited by M. A. Thalbourne and L. Storm, Jefferson, NC: McFarland, pp. 90–123.

Ertel, S. (2010). Psi in a skeptic's lab: A successful replication of Ertel's Ball Selection Test. *Journal of Scientific Exploration*, *14*(3), 461–478.

Ertel, S. (2013). Psi effect or sensory leakage: Scrutinizing the ball selection test. *Journal of Scientific Exploration*, *27*(3), 387–391.

Hong, S. M., & Faedda, S. (1996). Refinement of the Hong Psychological Reactance Scale. *Educational and Psychological Measurement*, *56*, 173–182.

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

Irwin, H. J. (2009). *The Psychology of Paranormal Belief. A Researcher's Handbook.* Hatfield, Herts, UK: University of Hertfordshire.

Kraus, S. J. (1995). *Attitudes* and the prediction of *behavior*: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, *21*, 58–75.

Lange, R., & Thalbourne, M. A. (2002). Rasch scaling paranormal belief and experience: The structure and semantics of Thalbourne's Australian Sheep–Goat Scale. *Psychological Reports*, *91*, 1065–1073.

Lawrence, T. R. (1990–1991). Subjective random generations and the reversed sheep–goat effect: A failure to replicate. *European Journal of Parapsychology*, *8*, 131–144.

Lawrence, T. R. (1993). Gathering in the sheep and goats: A meta-analysis of forced-choice sheep–goat ESP studies, 1947–1993. *Proceedings of the 36th Annual Convention of the Parapsychological Association*, pp. 75–86.

Lovitts, B. E. (1981). The sheep–goat effect turned upside down. *Journal of Parapsychology*, *45*, 293–309.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Services.

Miller, C. H., Burgoon, M., Grandpre, J., & Alvaro, E. (2006). Identifying principal risk factors for the initiation of adolescent smoking behaviors: The significance of psychological reactance. *Health Communication*, *19*, 241–252.

Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M. & Potts, K. A. (2007). Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom. *Human Communication Research*, *33*, 219–240.

Palmer, J. (1971). Scoring in ESP tests as a function of belief in ESP: Part I. The sheep–goat effect. *Journal of the American Society for Psychical Research*, *65*, 373–408.

Palmer, J. (1972). Scoring in ESP tests as a function of belief in ESP: Part II. Beyond the sheep–goat effect. *Journal of the American Society for Psychical Research*, *66*, 1–25.

Palmer, J. (1977). Attitudes and personality traits in experimental ESP research. In *Handbook of Parapsychology*, edited by B. B. Wolman, New York: Van Nostrand Reinhold, pp. 175–201.

Schmeidler, G. R. (1943). Predicting good and bad scores in a clairvoyance experiment: A preliminary report. *Journal of the American Society for Psychical Research*, *37*, 103–110.

Schmeidler, G. R. (1945). Separating the sheep from the goats. *Journal of the American Society for Psychical Research*, *39*, 47–49.

Schmeidler, G. R. (1988). *Parapsychology and Psychology: Matches and Mismatches.* Jefferson, NC: McFarland.

Schmeidler, G. R., & McConnell, R. A. (1973). *ESP and Personality Patterns.* Westport, CT: Greenwood.

Silvia, P. J. (2005). Deflecting reactance: The role of similarity in increasing compliance and reducing resistance. *Basic and Applied Social Psychology*, *27*, 227–284.

Silvia, P. J. (2006). Reactance and the dynamics of disagreement: Multiple paths from threatened freedom. *European Journal of Social Psychology*, *36*, 673–685.

Smith, M. J. (1978). Discrepancy and the importance of *attitudinal freedom*. *Human Communication Research*, *4*(4), 308–314.

Stanford, R. G. (1977). Conceptual frameworks of contemporary psi research. In *Handbook of Parapsychology* edited by B. B. Wolman, New York: Van Nostrand Reinhold, pp. 823–858.

Stanford, R. G. (1986). Experimental psychokinesis: A review from diverse perspectives. In *Handbook of Parapsychology* edited by B. B. Wolman, L. Dale, G. Schmeidler, and M. Ullman, Jefferson, NC: McFarland, pp. 324–381. [Original work published in 1977, New York: Van Nostrand Reinhold]

Steinkamp, F. (2005). Forced-Choice ESP Experiments: Their Past and Their Future. In *Parapsychology in the 21st Century: Essays on the Future of Psychical Research* edited by M. A. Thalbourne & L. Storm, Jefferson, NC: McFarland, pp.124–163.

Storm, L., & Thalbourne, M. A. (2005). The effect of a change in pro attitude on paranormal performance: A pilot study using naïve and sophisticated skeptics. *Journal of Scientific Exploration*, *19*, 11–29.

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., & Delin, P. S. (1994). A common thread underlying belief in the paranormal, creative personality, mystical experience, and psychopathology. *Journal of Parapsychology, 58*, 3–38.

Worchel, S., & Brehm, J. W. (1970). Effects of threats to attitudinal freedom as a function of agreement with the communicator. *Journal of Personality and Social Psychology*, *14*, 18–22.

Wright, R. A. (1986). Attitude change as a function of threat to attitudinal freedom and extent of agreement with a communicator. *European Journal of Social Psychology*, *16*, 43–50.