

The Sheep-Goat Effect

The term ‘sheep-goat effect’ was coined by [Gertrude Schmeidler](#) (1913-2009), professor of psychology at City University of New York. Schmeidler 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’).^[1] The definition has been extended to include sheep as those who ‘believe that ESP exists as a genuine phenomenon’,^[2] thus excluding goats from this belief. The sheep-goat effect refers to the significant paranormal (‘psi’) performance difference between sheep and goats, whereby sheep tend to perform well in psi tasks, scoring above mean chance expectation (MCE), whereas goats tend to perform poorly in psi tasks, scoring at or below MCE.

Outline of the Sheep-Goat Effect

Probably the best-known process-oriented variable that has been examined in relation to paranormal (‘psi’) performance is the sheep-goat variable, which is usually measured using paranormal belief scales or questions. Schmeidler was one of the first researchers who sought to measure participants’ beliefs about their psi performance, arguing that psi performance is related to paranormal belief.^[3] The *sheep-goat effect* is now taken to be ‘any significant psi performance scoring difference between these two groups [sheep and goats] as defined by the experimenter’.^[4]

To measure paranormal belief, Schmeidler started with simple questions, such as ‘Do you accept the possibility of ESP under the conditions of the experiment?’, to gauge whether a participant thought ESP might occur in an ESP experiment. For Schmeidler, participants’ response scores would be the means by which the experimenter could separate the sheep from the goats.^[5]

Confusion can arise, since sheep have come to be referred to as those who merely accept the possibility of ESP, and goats as those who do not accept the possibility of ESP. Therefore, high-scorers on paranormal belief scales are nominally labelled ‘sheep’, and low-scorers are labelled ‘goats’, but these classifications do not indicate actual level of psi performance (statistical proof of psychic ability).^[6] Nevertheless, the expectation is that sheep will tend to perform well in psi tasks, scoring above mean chance expectation (MCE), whereas goats will tend to perform poorly in psi tasks, scoring at or below MCE.

Statistically significant scoring above MCE is usually referred to as ‘psi-hitting’, and statistically significant scoring below MCE is referred to as ‘psi-missing’—both types of scoring are taken as proof of psi. Psi-hitting is assumed to be the product of attempts on the part of sheep to ‘prove’ the psi hypothesis, whereas psi-missing results from goats attempting to ‘disprove’ the same hypothesis.^[7] This principle is captured in [John Palmer](#)’s ‘vindication’ theory—the ‘need for vindication’ is a ‘need to defend the validity of one’s previously formed opinions on important social, moral, and political issues’.^[8]

Grounds for the expectation of a sheep-goat effect in psi experiments came early when Schmeidler and McConnell determined that about 80 percent of sheep-goat studies produce effects indicating that sheep score higher than goats.^[9] This expectation, supported by other studies (see next section), has led to a general understanding that the sheep-goat effect (SGE) is one of the more demonstrable forms of psi.^[10]

Empirical Findings

A sizeable number of tests of the SGE have been conducted, but there are only a few literature reviews on the effect,^[11] Tony Lawrence^[12] created a meta-analysis of the ‘forced-choice’ ESP literature from 1947 to 1993, while Storm and Tressoldi carried out a similar study in 2017.^[13] A meta-analysis of the ‘free-response’ literature has never been published. The typical forced-choice task requires participants to identify concealed targets such as symbols, letters, shapes, numbers, and so on. That is, the target-guess is ‘one of a limited range of possibilities which are known to [the participant] in advance’.^[14] Forced-choice differs from free-response in that the latter ‘describes any test of ESP in which the range of possible targets is relatively unlimited and is unknown to the percipient’.^[15]

The literature provides two reviews supportive of the SGE in the ESP domain, although neither is recent. First, Palmer presented an analysis of the then-current SGE literature (studies dating from 1947 to 1970). He found 13 out of 17 experiments (76%) were in the predicted direction. Six out of the 17 (35%) produced ‘significant confirmations of the sheep-goat hypothesis’.^[16]

Second, Palmer reported seven new experiments since his 1971 study, where five were confirmed to have produced significant effects in the direction hypothesized, and 6 out of 7 (86%) were in the predicted direction, though not necessarily significant.^[17] In total, Palmer’s two sets of studies combined give about 11 studies out of 24 (46%) that were significant and in the direction hypothesized, and 19 out of 24 (79%) that have produced effects in the hypothesized direction, though not necessarily significant. Palmer later found that ‘all the significant sheep-goat differences have been in the predicted direction’^[18] for studies where paranormal belief was confined to ESP ‘*in the test situation*’ (at a concrete level). There was further support for the SGE, although not as strong, in studies in which belief was measured as belief in ESP generally, at an abstract level, not just in the test situation.^[19]

In the 1990s, Lawrence conducted a meta-analysis on forced-choice ESP studies. He accumulated 73 studies (4,500 participants, 685,000 guesses), and calculated an SGE of 0.029, with a highly significant Stouffer Z of 8.17 ($p = 1.33 \times 10^{-16}$).^[20] Eighteen studies (24%) showed a significant SGE ($p = .05$). The mean z was 0.96, and mean SGE per investigator was 0.026. He also found that study quality and SGE had not changed in 46 years. The file-drawer estimate was 1,726 (23 unreported, nonsignificant studies for every one successful study). Lawrence concluded that there was a ‘belief-moderated communications anomaly’—in short, a sheep-goat effect.^[21]

It is important to note Steinkamp's critical observations of Lawrence's meta-analysis. 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)'.^[22] She also pointed out that it is not entirely clear whether the SGE holds for both naïve and experienced experimental participants, and hence it is not entirely clear what the sheep-goat variable really measures'.^[23]

The more recent meta-analysis by Storm and Tressoldi continues from where Lawrence left off, covering 49 studies carried out by 43 investigators between 1994 and 2015. Its findings were 'generally comparable' to Lawrence's, they write, reporting as follows:

The mean ES for ESP = .045, mean $z = 0.75$, Stouffer $Z = 5.23$ ($p = 8.47 \times 10^{-8}$), and the mean trial-based SGE = 0.034, mean $z = 0.24$, Stouffer $Z = 1.67$ ($p = .047$). ... The SGE did not vary significantly with belief measure used. Bayesian analysis of the same dataset yielded results supporting the 'frequentist' finding that the null hypothesis should be rejected.^[24]

They concluded that a 'belief-moderated communications anomaly' has been revealed in the forced-choice ESP domain 'that has been effectively uninterrupted and consistent for almost 70 years'.^[25]

Recent forced-choice studies have replicated the SGE.^[26] Smith and colleagues used Palmer's Criterion 4 ('expectations of how well one will do at the psi task')^[27] as their sheep-goat measure (for a description of all four criteria, see next section).^[28] Using an on-screen coin-flipper, the prediction of psi performance correlated positively and significantly with 'coin-flips correctly guessed'.^[29] However, other recent forced-choice studies failed to find an SGE.^[30] Nevertheless, in a relatively recent summary of the evidence for a number of potential psi-relevant variables in the forced-choice domain, Steinkamp concludes that there is 'partial evidence' that the SGE can be demonstrated by sheep who believe they can 'show ESP under experimental conditions',^[31] and 'promising evidence' that 'goats score low'.^[32]

The SGE has been investigated for the free-response domain to a limited degree. Parker reported an SGE in his review of a small number of 'ganzfeld' studies.^[33] An SGE was also found in a more recent ganzfeld study by Marcusson-Clavertz and [Etzel Cardena](#).^[34] However, other ganzfeld studies have failed to find SGEs.^[35] As implied above, a comprehensive meta-analytic review of free-response studies (or at least ganzfeld studies) would be a step towards generalizing the SGE to other experimental domains.

Belief Criteria and Attitudes towards ESP

Since Schmeidler's time,^[36] the scope of the sheep-goat variable has been considerably widened, to the point where it often refers to paranormal belief or disbelief in the abstract sense (a point raised in the previous section). In fact four distinct meanings of the term *sheep-goat variable* have been identified—these are referred to as 'criteria': (1) the participant believes ESP is possible during an ESP

experiment; (2) the participant believes in ESP in the abstract or theoretical sense; (3) the participant personally believes they have psychic ability; and (4) the participant believes he/she can or has scored above-chance in an ESP test.^[37] There is some evidence that these criteria may be a refined means of measuring the SGE, but Criteria 1 and 2 may be more helpful than Criteria 3 and 4, both of which may be 'less effective'.^[38] However, Lawrence found 'no overall relationship between type of belief measure on effect size', and no evidence that Criterion 1 was superior to the other three.^[39]

Palmer drew a distinction between *beliefs* about the existence of ESP, and *attitudes* towards it (i.e., 'whether the subject would *like* for ESP to exist').^[40] Despite the fact that there was no evidence at that point in time—that attitudes towards ESP correlate significantly with ESP scores—the issue was raised again by Irwin and Watt, who point out that attitude to ESP is a complex issue. Indeed, they go as far as saying that the SGE 'may stem from an attitude'.^[41] A study by Lovitts demonstrates this possibility.^[42] She divided participants into two groups—one comprised of 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 sheep appeared to have been manipulated to score like goats, and *vice versa*. While Irwin and Watt called for replication, they do suggest that the SGE may be a 'cognitive rather than motivational bias'.^[43] In other words, cognitive processes must underlie the motivation to try hard (or not try hard), but it is perhaps possible that other cognitive processes, triggered by various experimental treatments, may usurp prior motivations so that the participant may be put in the position of adopting, or at least considering, a new attitude and/or strategy that may change psi performance. Apart from the Lovitts study, other evidence suggests this may be the case.

Lawrence attempted a replication of Lovitts' experiment by expressly telling participants that the test was designed to prove ESP, or disprove ESP, depending on random assignment.^[44] Lawrence did not replicate Lovitts' interaction effect, but he did find a significant difference between test situation, especially in goats. Lawrence's study may have been compromised by the fact that there were no male sheep in the 'disprove-ESP' group, and no female goats in the 'prove-ESP' group. Lovitts and Lawrence both found that sex correlated significantly with belief, indicating that in Lawrence's study the 'prove-ESP' group did not sufficiently represent goats, and the 'disprove-ESP' group did not sufficiently represent sheep.

Evidence that goats can be manipulated into changing their psi performance comes from a study by Storm and Thalbourne.^[45] Their objective was to see if goats who were naïve about statistical inference could shift from chance-scoring (or psi-missing) to psi-hitting after having the implications of significance testing explained to them. The hypothesis was supported—in a symbol identification task, goats shifted from chance scoring (20%, where $P_{MCE} = 20\%$), to psi-hitting (30%, $p = .047$).

Walsh and Moddel conducted a clairvoyance task using Zener cards to examine the motivational role of belief.^[46] Participants were presented with written and verbal

statements of scientific data which either strengthened or challenged their beliefs. Believers given pro-psi statements performed significantly above MCE, and significantly better than groups who were non-believers or who received anti-psi statements. The authors concluded that ‘successful psi performance results from belief in psi, and not the reverse’.^[47] While they refer to belief reinforcement (a motivational change) to account for the best performance by pro-psi believers, they also refer to a ‘shifting’ ‘mind-set’, which could be regarded as attitudinal and cognitive change.^[48]

More recently, a different approach has been taken whereby change in attitude towards psi is couched in terms to do with psychological reactance. Four forced-choice studies have been conducted that directly test the manipulation of reactance, especially in goats.^[49] According to Reactance Theory,^[50] an individual’s freedom, if threatened by coercion (a form of reactance treatment), may result in reactance, which is ‘a motivational state aimed at restoring the threatened freedom’.^[51]

To test reactance theory a reactance treatment, or prime, in the form of an opinionated communication, is used. According to theory, the treatment raises reactance, which remains high if no outlet is provided. Since increased noncompliant behaviour in participants (especially goats) is expected when they are under threat, increased target *avoidance* and therefore shifts from chance scoring to psi missing may follow.

Using a Ball-Selection Test,^[52] involving long-run predictions of numbered ping-pong balls drawn blindly from a black bag, psi scoring was found to be significantly lower in sheep and goats who were reactance primed compared to controls who were not.^[53] Also, the difference between reactant goats and control sheep was significant in two tests (one on the full dataset, and the other on the first-run dataset only).

It has also been found that the reactance treatment is detrimental to psi performance (especially in goats) in forced-choice experiments involving the Chinese system of divination, the *I Ching*.^[54] In two studies, participants graphically represented their cognitive/emotional states using an inverted pyramid-shaped Q-Sort Grid to rank 64 *I Ching* descriptor-pairs from -7 to +7. They then used a random number generator to generate an *I Ching* hexagram with an associated reading. The higher the hexagram ranking, referred to as a Q-Sort score, the better (more accurate) the prediction. The mean Q-Sort score of the reactance-treated group was lower than the control group, although the result did not reach significance. As expected, however, the mean Q-Sort score of reactance-treated goats was significantly lower than control goats. A study using Zener cards also yielded a similar directional finding for sheep and goats, although the difference was not significant.^[55]

In summary, Palmer made the point decades ago that paranormal belief can be measured at simple (concrete) and abstract levels.^[56] Collectively, experimenters seem to be attending to this dual aspect in their designs by using a mixed-bag of assessment tools—psychometrically sound multiple-item scales, short-scales (such as two or three questions), and even single questions. As a consequence, while it

has come down to us from Lawrence that the SGE is 'robust', there being 'no single best way of separating sheep from goats',^[57] the work to find the best, most reliable belief predictor(s) of psi performance is ongoing.

While paranormal belief predicts ESP performance adequately, and seemingly regardless of the belief measure used, a small number of studies have so far supported the claim that psi performance may change if attitude (mindset) is changed. In controlled studies where direct causal assumptions can be made, it would appear that various kinds of treatment (e.g., alternative instructions, reactance priming) can be instrumental in eliciting attitude-related psi performances that are superior or inferior to prior performances. These findings suggest that ESP performance is not an entrenched motivational response driven exclusively by an immutable paranormal belief, but is driven by cognitive biases.

Conclusion

The sheep-goat effect refers to any significant psi performance scoring difference between so-called sheep and goats, where these two terms 'sheep' and 'goats' are defined by the given experimenter, but usually refer to psi believers and psi non-believers, respectively. The terms sheep and goat may often be 'nominal' only, having been assigned on the basis of a score on a paranormal belief measure, without actually referring to psi performance in an ESP task.

Based on pivotal studies,^[58] paranormal belief, as measured on sheep-goat scales or individual questions, tends to be a predictor of psi outcomes, with sheep producing above-chance hit rates, and goats producing hit-rates at, or below-chance.

The sheep-goat variable can be defined by four belief criteria, which indicate the subtle interpretative differences underlying paranormal belief, but evidence thus far suggests that they all predict the sheep-goat effect to the same degree.

There is some evidence suggesting that attitudes of sheep and goats towards psi can be changed through various treatments, and these changes can elicit changes in psi performance. Thus, the sheep-goat effect may be underscored more by attitude than motivation (motivational bias), suggesting that cognition (cognitive bias) plays a defining role in the sheep-goat effect.

Lance Storm

Literature

Billows, H. & Storm, L. (2015). Believe it or not: A confirmatory study on predictors of paranormal belief, and a psi test. *Australian Journal of Parapsychology* 15, 7-35.

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

Broughton, R.S., & Alexander, C.H. (1997). Autoganzfeld II: An attempted replication of the PRL ganzfeld research. *Journal of Parapsychology* 61, 209-26.

Cardeña, E., Marcusson-Clavertz, D. & Wasmuth, J. (2009). Hypnotizability and dissociation as predictors of performance in a precognition task: A pilot

study. *Journal of Parapsychology* 73, 137-58.

Ertel, S. (2005). The ball drawing test: Psi from untrodden ground. In *Parapsychology in the Twenty-First Century: Essays on the Future of Psychical Research*, ed. by M.A. Thalbourne & L. Storm, 90-123. Jefferson NC: McFarland.

Hitchman, G.A.M., 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* 76/1, 109-45.

Irwin, H.J. (1993). Belief in the paranormal: A review of the empirical literature. *Journal of the American Society for Psychical Research* 87, 1-39.

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

Lawrence, T.R. (1990-91). Subjective random generations and the reversed sheep-goat effect: A failure to replicate. *European Journal of Parapsychology* 8, 131-44.

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

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

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 Society for Psychical Research* 72, 193-207.

Marcusson-Clavertz, D. & Cardena, E. (2011). Hypnotizability, alterations in consciousness and other variables as predictors of performance in a ganzfeld psi task. *Journal of Parapsychology* 75, 235-59.

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*, ed. by B.B. Wolman, 175-201. New York: Van Nostrand Reinhold.

Palmer, J. (1978). Extra-Sensory Perception, research findings. In *Advances in Parapsychological Research Volume 2*, ed. by S. Krippner, 59-243. New York: Plenum Press.

Parker, A. (2000). A review of the ganzfeld work at Gothenburg University. *Journal of the Society for Psychical Research* 64, 1-15.

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, 103-10.

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

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

Schönwetter, T., Ambach, W. & Vaitl, D. (2011). Does a modified guilty knowledge test reveal anomalous interactions within pairs of participants? *Journal of Parapsychology* 75, 93-118.

Silvia, P.J. (2005). Deflecting reactance: The role of similarity in increasing compliance and reducing resistance. *Basic and Applied Social Psychology* 27, 227-84.

Smith, M.D., Wiseman, R., Machin, D., Harris, P. & Joiner, R. (1997). Luckiness, competition and performance on a psi task. *Journal of Parapsychology* 61, 33-43.

Steinkamp, F. (2005). Forced-choice ESP experiments: Their past and their future. In *Parapsychology in the Twenty-First Century: Essays on the Future of Psychical Research* ed. by M.A. Thalbourne & L. Storm, 124-63. Jefferson, NC: McFarland.

Storm, L., & Tressoldi, P.E. (2017). Gathering in more sheep and goats: A meta-analysis of forced-choice sheep-goat studies, 1994-2015. *Journal of the Society for Psychical Research* 81/2, 79-107.

Storm, L., Ertel, S. & Rock, A.J. (2013). Paranormal effects and behavioural characteristics of participants in a forced-choice psi task: Ertel's Ball 'Selection Test' under scrutiny. *Australian Journal of Parapsychology* 13, 111-31.

Storm, L. & Rock, A.J. (2014). An investigation of the *I Ching* using the Q-Sort Method and an RNG-PK design: II. The effect of reactance on psi. *Australian Journal of Parapsychology* 14, 163-90.

Storm, L., & Thalbourne, M.A. (2005). The effect of a change in pro attitude on paranormal performance: A pilot study using naive and sophisticated skeptics. *Journal of Scientific Exploration* 19, 1-29.

Thalbourne, M.A. (2003). *A Glossary of Terms Used in Parapsychology*. Charlottesville, VA: Puente.

Thalbourne, M.A. (2010). Transliminality: A fundamental mechanism in psychology and parapsychology. *Australian Journal of Parapsychology* 10, 70-81.

Walsh, K., & Moddel, G. (2007). Effect of belief on psi performance on a card guessing task. *Journal of Scientific Exploration* 21, 501-10.

Endnotes

Footnotes

1. ^ Schmeidler (1943, 1945).
2. ^ Thalbourne (2003), 114.
3. ^ Schmeidler (1943, 1945).
4. ^ Thalbourne (2003), 114.
5. ^ The terms 'sheep' and 'goat' were adopted by Schmeidler from a New Testament simile that describes how a shepherd 'separates the sheep from the goats' (Matthew 25: 31-33).
6. ^ Schmeidler & McConnell (1973).
7. ^ Palmer (1971; 1972); Schmeidler & McConnell (1973).
8. ^ Palmer (1972), 10.
9. ^ Schmeidler & McConnell (1973).
10. ^ For historical examples of empirical evidence for the sheep-goat effect, see Lawrence (1993), Palmer (1977).
11. ^ Irwin (1993); Palmer (1971; 1972; 1977); Schmeidler & McConnell (1973); Thalbourne (2010).
12. ^ Lawrence (1993).
13. ^ Storm & Tressoldi (2017).
14. ^ Thalbourne (2003), 44.
15. ^ Thalbourne (2003), 44.
16. ^ Palmer (1971), 402.
17. ^ Palmer (1977).
18. ^ Palmer (1978), 154.
19. ^ Palmer (1978), 155.
20. ^ Lawrence (1993).
21. ^ Lawrence (1993), 75.
22. ^ Steinkamp (2005), 152-53.
23. ^ Steinkamp (2005), 153.
24. ^ Storm & Tressoldi (2017), 79.
25. ^ Storm & Tressoldi (2017), 79.
26. ^ For example, Luke et al. (2008); Smith et al. (1997).
27. ^ Palmer (1971), 394-95.
28. ^ Smith et al. (1997), 40.
29. ^ Smith et al. (1997), 39.
30. ^ Cardena et al. (2009); Hitchman et al. (2012); Schönwetter et al. (2011).
31. ^ Steinkamp (2005), 153.
32. ^ Steinkamp (2005), 156.
33. ^ Parker (2000). The ganzfeld design is a type of free-response experiment which uses relaxation techniques and sensory (visual and auditory) homogenization to help facilitate a physical and mental environment conducive to psi.
34. ^ Marcusson-Clavertz & Cardena (2011).
35. ^ Broughton & Alexander, (1997); Cardena et al. (2009); Hitchman et al. (2012); Schönwetter et al. (2011).
36. ^ Schmeidler (1943).
37. ^ Palmer (1971), 391-94.
38. ^ Palmer (1971), 396.
39. ^ Lawrence (1993), 80.
40. ^ Palmer (1978), 160.

41. ^ Irwin & Watt (2007), 75.
42. ^ Lovitts (1981).
43. ^ Irwin & Watt (2007), 75.
44. ^ Lawrence (1990-91).
45. ^ Storm & Thalbourne (2005).
46. ^ Walsh & Moddel (2007).
47. ^ Walsh & Moddel (2007), 501.
48. ^ Walsh & Moddel (2007), 505.
49. ^ See Billows & Storm (2015); Storm (2016); Storm et al. (2013); Storm & Rock (2014).
50. ^ Brehm (1966).
51. ^ Silvia (2005), 277
52. ^ Ertel (2005).
53. ^ Storm et al. (2013).
54. ^ Storm & Rock (2014).
55. ^ Billows & Storm (2016).
56. ^ Palmer (1978).
57. ^ Lawrence (1993), 81.
58. ^ Lawrence (1993); Palmer (1971; 1977); Schmeidler & McConnell (1973).