# On the correspondence between dream content and target material under laboratory conditions: A meta-analysis of dream-ESP studies, 1966-2016

Lance Storm[1] Simon J. Sherwood[2], Chris A. Roe[3], Patrizio E. Tressoldi[4], Adam J. Rock[5], & Lorenzo Di Risio[6]

[1]School of Psychology, University of Adelaide, Adelaide, Australia.
[2]College of Life and Natural Sciences (Psychology), University of Derby, Derby, UK.
[3]Division of Psychology, University of Northampton, Northampton, UK.
[4]Dipartimento di Psicologia Generale, Università di Padova, Padua, Italy.
[5]School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale NSW, Australia.
[6]Dipartimento di Psicologia Generale, Università di Padova, Padua, Italy.

*Summary*. In order to further our understanding about the limits of human consciousness and the dream state, we report meta-analytic results on experimental dream-ESP studies for the period 1966 to 2016. Dream-ESP can be defined as a form of extra-sensory perception (ESP) in which a dreaming perceiver ostensibly gains information about a randomly selected target without using the normal sensory modalities or logical inference. Studies fell into two categories: the Maimonides Dream Lab (MDL) studies ($n$ = 14), and independent (non-MDL) studies ($n$ = 36). The MDL dataset yielded mean $ES$ = .33 ($SD$ = 0.37); the non-MDL studies yielded mean $ES$ = .14 ($SD$ = 0.27). The difference between the two mean values was not significant. A homogeneous dataset ($N$ = 50) yielded a mean $z$ of 0.75 ($ES$ = .20, $SD$ = 0.31), with corresponding significant Stouffer $Z = 5.32$, $p = 5.19 \times 10^{-8}$, suggesting that dream content can be used to identify target materials correctly and more often than would be expected by chance. No significant differences were found between: (a) three modes of ESP (telepathy, clairvoyance, precognition), (b) senders, (c) perceivers, or (d) REM/non-REM monitoring. The ES difference between dynamic targets (e.g., movie-film) and static (e.g., photographs) targets was not significant. We also found that significant improvements in the quality of the studies was not related to $ES$, but $ES$ did decline over the 51-year period. Bayesian analysis of the same homogeneous dataset yielded results supporting the 'frequentist' finding that the null hypothesis should be rejected. We conclude that the dream-ESP paradigm in parapsychology is worthy of continued investigation, but we recommend design improvements.

*Keywords:* Bayesian analysis, dream ESP, ESP, meta-analysis, paranormal, psi

## 1. Introduction

For more than half a century, considerable research has been conducted into the alleged occurrence of extrasensory perception (ESP) in dreams (for reviews, see Roe & Sherwood, 2009; Sherwood & Roe, 2003, 2013; Van de Castle, 1977, 2009). For the purposes of clarification, one reviewer of this paper defines the typical dream as "a series of images, thoughts, and/or feelings that occur during sleep, and which often can be recalled and reported upon awakening.". ESP refers to the apparent acquisition of information about the environment that seems not to be mediated by the usual sensory modalities, or by inference from information that is conventionally available. Instances of dream ESP can be categorised as *telepathy*, *clairvoyance*, and *precognition*. Telepathy refers to the "paranormal acquisition of information concerning the thoughts, feelings or activity of another conscious being" (Thalbourne, 2003, p. 125). Clairvoyance is defined as "paranormal acquisition of information concerning an object or contemporary physical event; in contrast to telepathy, the information is assumed to derive directly from an external physical source" (Thalbourne, 2003, p. 18). Precognition is defined as "a form of extrasensory perception in which the target is some future event that cannot be deduced from normally known data in the present" (Thalbourne, 2003, p. 90). In practice these categories may be difficult to distinguish and tend to be defined operationally.

Research into ostensible 'psi' during the dream state includes surveys and spontaneous case collections (e.g., Steinkamp, 2000). Thouless and Wiesner (1947) proposed

Corresponding address:
Lance Storm, PhD. School of Psychology, University of Adelaide, Australia.
Email: lance.storm@adelaide.edu.au

Submitted for publication: January 2017
Accepted for publication: July 2017

a unitary process underlying ESP and PK that they labelled 'psi' (from the Greek letter ψ), and can be used as an adjective or noun "to describe paranormal processes and paranormal causation" (Thalbourne, 2003, p. 92). However, it is beyond the scope of this meta-analytic paper to conduct a review of broader domains of dream-psi studies. Instead, we necessarily restrict our review to dream-ESP studies that have taken place in the laboratory, or were conducted under similarly controlled 'non-laboratory' (home) experimental conditions utilizing, in both cases, randomly selected target sets. Considerable effort is required to operationalize a dream-ESP experiment, so it is perhaps not surprising that dream-ESP research has waned in recent decades. However, in the interests of the scientific enterprise, we aim to determine if the outcomes and findings of dream-ESP research justify (and have justified) the human effort and financial costs necessary to conduct this kind of research, particularly given the controversial status of claims for ESP effects.

Thalbourne (2003) does not define the term 'dream ESP', but he does refer to the so-called *veridical dream*, which is "an apparently paranormal dream, inasmuch as some of the dream details give information about events normally unknowable to the experient" (p. 33). For our purposes, dream ESP involves ostensibly paranormal communication while in an altered state of consciousness (ASC) commonly known as dreaming. According to Krippner (1972) an ASC may be defined as "a mental state which can be subjectively recognized by an individual (or by an objective observer of the individual) as representing a difference in psychological functioning from the individual's 'normal' alert state" (p. 1). (Rock and Krippner, 2012, later stated that this definition "does not state whether the pattern and/or intensity of the psychological functioning must be different compared to the percipient's 'normal alert state' for an ASC to occur", pp. 6-7).

The specific form of ASC, the *dream state*, is considered particularly conducive to psi because the field of consciousness is reduced—in a strong sense it resembles the state elicited in the laboratory treatment known as the *Ganzfeld* ("total field"; see Storm, Tressoldi, & Di Risio, 2010) because stimulation from all the sensory modalities is considerably reduced, or even blocked completely. Specifically, the Ganzfeld—a "homogeneous perceptual environment" (Bem, 1993, p. 102)—consists of an undifferentiated visual field created by viewing a red light through halved table-tennis balls taped over a percipient's eyes. Additionally, an analogous auditory field is produced by listening to stereophonic white or pink noise (i.e., a monotonous hissing sound; Bem, 1993). Like the Ganzfeld state, the dream state thus may enable any 'psi signal' the best possible chance of being detected above sensory noise. We note, however, that 'informational' concepts such as *psi signal* and *sensory noise* are disputed in some parapsychological circles (see Jung, 1960; Stanford, 1977, 1978).

Like the Ganzfeld design, the typical dream-ESP experiment requires a 'sender' (the one who 'sends' or transmits the target image) and a dreaming 'receiver' (the one who 'receives' or accepts the target image). We note that these days the terms 'agent' and 'percipient' (or perceiver), respectively, are preferred by some parapsychologists who question the assumption that ESP involves information exchange, as it may merely be a "correlation in an entangled psycho-physical system" (von Lucadou, 2001, p. 13). Either way, psi is seen to be encapsulated in the dream process, with the psi target seemingly embedded in the imagery that is the dream content.

Dreaming is a convenient ASC upon which researchers may focus because such states are naturally occurring, and therefore require no special training or induction procedures. Dreaming also has features considered to be important facilitators of ESP (see Braud & Braud, 1975; Honorton, 1977; Parker, 1975). However, dreaming has the disadvantage of requiring expensive equipment so that brain activity may be monitored if participants are to be deliberately awakened from their dreams in order to report them. Such monitoring of brain activity takes the usual forms of the electroencephalogram (EEG) and eye movement or electro-oculogram (EOG) that are indicative of the rapid-eye-movement dream-stage of sleep. Rapid-eye-movement (REM) sleep is characterised by the appearance of phasic bursts of rapid jerky eye movements, reports of dreaming if woken, a highly active brain similar to that in the waking state, and muscular paralysis (Dement & Kleitman, 1957; Rechtschaffen, 1973). Laboratory studies have shown that, on average, dreams are reported on about 80% of awakenings from REM sleep (see Goodenough, 1991).

One of the aims of the present study is to test for an effect-size difference between REM and non-REM dream-ESP studies to see if a REM-psi relationship can be substantiated experimentally. While the REM-state is associated with dream activity, any differences between the two types of dream-state (REM and non-REM) may indicate whether REM activity itself plays a crucial role in the dream/ESP relationship. More broadly, in this paper we will evaluate a comprehensive database of dream-ESP studies conducted over the period 1966 to 2016 inclusive, and we will compare the earlier so-called Maimonides studies with independent (non-Maimonides) studies. These studies all use dream reports as the data to be assessed. We will attempt to identify possible reasons for any differences in alleged ESP performances brought about by various conditions (described below). We will conclude by making some recommendations for future research of this type.

Before we begin our analysis, some background information about the innovative early work conducted at the Maimonides Medical Center and subsequent (mainly conceptual) replications will illustrate the typical protocols adopted in dream-ESP studies, as well as give a context for our study, and clarify some important points not yet raised.

### 1.1. The Maimonides Dream-ESP Studies

Case collections bear out the suggestion that a large proportion of spontaneous occurrences of alleged ESP manifest during dreams—between 63% (Sannwald, 1963) and 65% (Rhine, 1981). There are also a number of reports of clients experiencing some form of dream ESP during therapy (e.g., see Krippner, 1991; Van de Castle, 1977). Such instances in a clinical context were sufficiently impressive for prominent psychoanalyst and psychiatrist Dr Montague Ullman to initiate some dream-ESP research in 1960 with Mrs Eileen Garrett, who was a well-known medium, supporter of psychical research, and founder of the Parapsychology Foundation. Having set up a temporary dream lab with psi researchers Dr Karlis Osis and Douglas Dean, Ullman (1969) used three pictures from *Life* magazine as a target set. One of the three pictures—a still of the chariot race in the film *Ben Hur*—was telepathically sent to Garrett, who later described the pic-

ture accurately (for details about this incident, see Radin, 2006, p. 106).

Having had some success with Garrett using this method, Ullman set up the Maimonides Dream Laboratory (MDL) at the Maimonides Medical Center in Brooklyn in 1962 (Krippner, 1991; Ullman & Krippner with Vaughan, 1973, 1989, 2002). The 'Maimonides' procedure was developed and improved over time and a number of procedural variations were utilized. The following, therefore, is a general description of the features of a trial designed to investigate dream telepathy.

The percipient was attached to EEG-EOG monitoring equipment and slept in a sound-attenuated room in the laboratory. Once he or she was asleep, a target was randomly selected from among a pool of targets (typically art prints), compiled on the basis of the images' emotional intensity, vividness, color, and simplicity.

The target, in a sealed envelope, was given to the sender, who was then locked inside another sound-attenuated room in the building (or, in some studies, a different building). The experimenter monitored the percipient's EEG-EOG throughout the night and, once this indicated that the percipient had entered REM sleep, signaled the agent (via a buzzer) to open the target envelope and begin 'sending' the target. At, or towards, the end of the REM period, the experimenter awakened the percipient via an intercom and asked him or her to describe any dream(s) they could recall. Responses throughout the night and in the morning were tape-recorded and later transcribed. The agent heard the percipient's dream report via a loudspeaker, which may have reinforced his or her subsequent sending strategy. The percipient then went back to sleep.

The above process was repeated for each REM period with the same target being sent each time. In the morning, the percipient reported any associations to the dream mentation and guessed what the target might be. Percipients typically viewed between eight and twelve pictures, one of which was the target, gave a confidence rating for each picture and also placed them in rank order according to the correspondence with their dream mentation, associations, and/or guesses. Complete dream transcripts and target sets were also sent to two or three independent judges who made similar judgments. The ratings/rankings from the blind judges were combined. A trial was a 'binary hit' if the target picture had been ranked in the top half of the target set and a 'binary miss' if ranked in the bottom half. Performance was then evaluated to determine whether it was significantly higher or lower than mean chance expectation (MCE). The following example is illustrative of the dream process, and how dream reports ostensibly correspond with a given random target (reported in Rao, 2001, p. 150):

> Hiroshige's "*Downpour at Shono*" was randomly selected. It portrays a Japanese man with an umbrella trying to escape a driving rain. The directions in the box of multisensory materials read "Take a shower." A small Oriental umbrella was included in the box.
>
> - *First Dream Report*: No apparent correspondences.
> - *Second Dream Report*: "It's as though I was doing some drawing, or some drawing was being done. This was very hazy ... I had the feeling as though it were in a *down position*, like a low table. *Down* on the floor. Seems that's what I meant by '*down*'."
> - *Third Dream Report*: "... Something about *an Oriental man* who was ill ..."
> - *Fourth Dream Report*: "*… it had to do with fountains—a big fountain*. It would be like one you see in Italy. *A fountain*. Two images and a *water spray* that would shoot up. No color."

Before the MDL closed in 1978, Ullman's research team had conducted 13 formal dream-ESP studies (including an incomplete study by Honorton, Ullman, & Krippner, 1975) and three groups of pilot sessions (Krippner, 1991, 1993; Ullman et al., 1973, 1989), thus yielding a total of 15 studies if we necessarily exclude Honorton et al. (1975). Of the 12 usable formal studies, 10 were designed to investigate telepathy, and two were designed to investigate precognition. The pilot sessions were designed to investigate telepathy, precognition, and clairvoyance, respectively. All 15 studies form part of our database (listed in Table A1 of Appendix A).

### 1.2. Other Dream-ESP Studies

During the 1960s and 1970s there were six independent/semi-independent replication attempts by researchers at other laboratories using EEG-EOG monitoring and deliberate awakening from REM sleep (Belvedere & Foulkes, 1971; Dement, 1974; Foulkes et al., 1972; Globus, Knapp, Skinner, & Healy, 1968; Hall, 1967; Strauch, 1970). However, five of these studies (i.e., Belvedere & Foulkes, 1971; Dement, 1974; Globus et al., 1968; Hall, 1967; Strauch, 1970) cannot be considered exact replication attempts because of variations in procedures. Also, some of the investigators in the Foulkes et al. (1972) study had been involved in a previous Maimonides study, thus compromising the independence of this replication attempt. Three studies are difficult to evaluate due to the limited amount of detail available in the published reports (Dement, 1974; Globus et al., 1968; Strauch, 1970), and one was in German (Hall, 1967).

The limited number of exact replications may be due to the prohibitive costs of maintaining a sleep laboratory. However, some researchers have been able to investigate dream ESP by developing less expensive and less labor-intensive methods. On this point, as Markwick and Beloff (1988) so succinctly put it, "Home dream research has much to contribute to this exciting field" (p. 81). Thus the majority of post-Maimonides studies involved the participants sleeping in their own homes rather than in a laboratory—participants who sleep at home are likely to feel more comfortable and so can awaken naturally and their sleep routines are less likely to be disrupted. Nevertheless, it is important to allow at least one pilot night for the participants to adjust to the study demands before the experiment begins, much as one would do with a dream-laboratory study.

The advantage, however, of awakening participants from REM sleep is that dream recall is much more likely, and can lead to more detailed and longer overall reports. Reviews of studies involving laboratory awakening from REM have concluded that dreams are reported in about 75-80% of cases (see Empson, 2002; Goodenough, 1991). Spontaneous awakenings in the morning are less likely to lead to dream recall, and any dreams that are reported tend to be those from the last REM period only (Empson, 2002). The Maimonides procedure tended to ask participants for their associations as well as their guesses, which is likely to have generated more—and richer—information upon which judges could base their judgments.

There are two other major differences between MDL and non-MDL studies as found by Sherwood and Roe (2003, 2013) in their narrative reviews. They (Sherwood & Roe, 2003) first identified 21 statistically assessable dream-ESP datasets that were conducted in the 25 years since the MDL closed. Later, they (Sherwood & Roe, 2013) found a further seven datasets, thus bringing the total post-Maimonides count to 28. Unlike the MDL series, which focused mainly on telepathy, less than half of the post-Maimonides studies did so. The majority investigated clairvoyance (i.e., did not include a sender), which may have been preferred because it is methodologically simpler and precludes some channels of normal communication. Also, the Maimonides studies featured targets that had emotional themes noted for their "vividness, colour, and simplicity" (Sherwood & Roe, 2013, p. 72), whereas post-Maimonides studies tended to use more neutral targets.

It is likely that these methodological differences produce subsequent dream-ESP performance differences, but tests so far are inconclusive: Sherwood and Roe (2013) found a difference when testing *judges*' scores from the MDL set against post-MDL set (the "Maimonides studies were significantly more successful than the post-Maimonides studies in terms of effect size", p. 67), whereas Storm and Rock (2015) found no significant difference when they tested *participants*' scores from the MDL set against the post-MDL set. The major aim of the present study, using a more up-to-date database, is to determine if there are MDL/non-MDL dream-ESP performance differences to gain a clearer picture of the achievements of the MDL and non-MDL databases. This update is necessary since both studies by Sherwood and Roe (2003, 2013) were narrative reviews (not meta-analyses), and their statistical findings are not entirely reliable due to missing data, or unsystematic criteria for the inclusion and exclusion of studies. For example, an unpublished dissertation (Eppinger, 2001), not peer-reviewed, was included in both reviews, while other studies were excluded from both reviews (Belvedere & Foulkes, 1971; Foulkes et al., 1972; Van de Castle, 1971). The present study thus more accurately reflects the dream ESP literature.

### 1.3. Differentiating Telepathy, Clairvoyance, and Precognition

It may seem unusual to characterise various ESP tests, including dream ESP, as tests of telepathy or clairvoyance given that the psi hypothesis effectively knows no bounds, so that there may be no valid means of ruling that a given study is indeed a study of telepathy and not clairvoyance, or *vice versa* (with precognition we may be on safer ground—see next paragraph). We can see that the issue cannot be resolved where a so-called 'mental' event (a thought) exists just as cogently as a target (about which that thought originates) that could be observed/perceived clairvoyantly. As Bauer (1984) points out (speaking on behalf of Rhine, 1974):

> *a large number of parapsychological research topics, such as . . . telepathy are basically insolvable problems which cannot be studied empirically as it is impossible to eliminate clairvoyance as a potential alternative hypothesis. It is safe to assume that this dilemma is not simply a semantic one. It reflects principally different theoretical models which have of course consequences for the empirical testability of the hypotheses derived therefrom. (p. 143)*

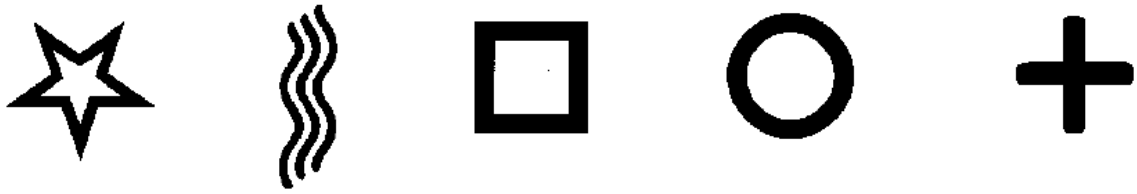
Perhaps telepathy, or at least a pure form of it, may even be an abstraction that is not only impossible to prove, but is parsimoniously redundant. Small wonder some researchers test an ambiguous form of telepathy/clairvoyance known as *general extra-sensory perception* (GESP), and avoid the theoretical issue altogether. Assuming there is a theoretical impasse, it may come as no surprise that Storm et al. (2010) found no significant effect-size difference between the three modalities (telepathy, clairvoyance, and precognition), and Storm, Tressoldi and Di Risio (2012) found no difference between telepathy and clairvoyance. Earlier, Steinkamp, Milton, and Morris (1998), using a total of 22 comparable pairs of precognition and clairvoyance studies (where procedures were effectively the same between types), found that effect sizes for precognition and clairvoyance were almost identical. These results either mean that ESP is a consistent effect across the three modalities, or one and the same ESP phenomenon is demonstrated each time across the three modalities which are merely provisional or working 'constructs' waiting for unification. Of course, there is a third option implied in Bauer's (1984) quote—some number of studies (if not all) are only nominally testing telepathy for clairvoyance, or *vice versa*.

While Bauer's (1984) words are sobering, there is an alternative view in which precognition could explain both telepathy and clairvoyance. In fact, it may be that a single precognitive theory would suffice in explaining many (if not all) forms of psi. For example, Decision Augmentation Theory (May, Utts, & Spottiswoode, 1995) explains ESP and some forms of psychokinesis (i.e., PK, which refers to paranormal mental influence on matter) as being efficacious information acquisition from the past, present, and future, incorporated unconsciously into the mind continuously so as to influence the organism's mental and behavioural decisions—hence, *decision augmentation*. There are other theoretical attempts at describing forms of ESP/PK unification including 'synchronicity' (Jung, 1960), 'psychopraxia' (Thalbourne, 2004), and possibly the 'Psi-Mediated Instrumental Response' model (Stanford, 1978), but space prohibits descriptions of these theories. Until such time as a new theory can better inform us, we will, in the present meta-analysis, compare the three major modes of ESP (telepathy, clairvoyance, and precognition) to see if psi effects vary across modalities.

### 1.4. Target Types

In ESP studies, target types have varied considerably over the decades. Examples include so-called 'static' targets such as face and symbol cards from the 52-card playing deck, Zener cards (Rhine et al., 1940/1966; see Figure 1), pictures (Watt, 1996), drawings (Simmonds-Moore & Moore, 2009), letters (Vaughan & Houck, 1993, 2000), numbers (Palmer, 2009), and so-called 'dynamic' targets such as movie and video clips (Honorton, Ullman, & Krippner, 1975; Parker, 2005). Other experimenters have opted for more ecologically valid targets based on the premise that these are more in keeping with real-world or divinatory scenarios where psi might therefore be better elicited. Examples include Roe, Davey, and Stevens (2003) with their horse racing design, Ivtzan and French (2004) with Tarot reading (see also, the review by Ivtzan, 2007), and Storm and Thalbourne (1998-1999) with their *I Ching* design. More recently, under the same premise, Ertel (2005) used physical targets tested at home (see also, Storm, Ertel, & Rock, 2013).

Figure 1. The Zener-card symbols: Star, Wavy Lines, Square, Circle, and Cross.



It is a moot point whether target types should be interesting or meaningful to participants on the assumption that emotional stimulation and/or meaningful targets (e.g., divinatory readings, real pictures, video clips) induce stronger psi effects. Indeed, shifts from significant to nonsignificant effects can be shown to occur within designs (see Storm's, 2009, *I Ching* series of studies) so that even if there is some degree of participant interest (i.e., emotional stimulation) in divinatory readings, interest does not necessarily lead to a significant psi effect. As it happens, Storm et al. (2012) found no difference between three target types: pictures/drawings, words/letters, and objects (i.e., targets that occupy 3-D physical space). However, in the Ganzfeld domain, Bem and Honorton (1994) found, as predicted, that the hit rate was significantly higher for dynamic than static targets (37% and 27% respectively, where MCE = 25%). This prediction was made in an earlier study by Honorton et al. (1990). In the present study, we consider the issue of target type (dynamic vs. static) and test the hypothesis that target type modifies effect size.

### 1.5. High *k*-Choice Designs

It is claimed that declines in effect size might be related to the number of target choices, *k*, in a target set. Timm (2000) argued that effect size measures have limited use if they do not adequately account for *k*. He argues that

> *the significance of ESP experiments must increase not only with N [the size of the sample or number of trials] but also with decreasing hit probability P (or with increasing number of target alternatives k = 1/P). (p. 253)*

Storm et al. (2012) found support for Timm's claim. They tested *z* scores across six levels of *k* ( 2, 3, 4, 5, 6, and 26), and found a significant correlation—*z* scores tended to increase as *k* increased, $r(5) = 0.79$, $p = .03$ (one-tailed). In the present study, we will again test *z* scores against *k* values.

### 1.6. Design of the Present Study

Notwithstanding the above reviews by Sherwood and Roe (2003, 2013), the dream-ESP database has never been formally meta-analysed. A comprehensive up-to-date meta-analysis of the dream-ESP domain (including planned hypotheses and statistical tests) is therefore long overdue with studies going back to 1966. Also, performance differences between Maimonides Dream Laboratory (MDL) studies and non-Maimonides (non-MDL) studies have not been conclusively tested. We also note that dream-ESP studies (MDL and non-MDL) tested: (a) same perceivers and different perceivers across dream trials; (b) same agents (i.e., senders) and different agents; (c) single perceivers and multiple perceivers; and (d) so-called star-subjects in single-subject studies (i.e., $N_p = 1$, where $N_p$ is number of percipients) and multiple-perceivers (i.e., $N_p > 1$). We point out that (a) and (c) do differ (i.e., "same perceiver" and "single perceiver" are not necessarily identical). On the one hand, 'same perceiver' in (a) can mean studies in which multiple perceivers worked *together* across the *same set of trials* for a consensus vote (e.g., Dalton et al., 1999; Dalton et al., 2000), but *not* studies (such as Foulkes et al., 1972; Kanthamani & Broughton, 1992) in which perceivers worked *separately*, each getting their *own unique* trial or set of trials. On the other hand, 'single perceiver' in (c) means a perceiver *did* work *separately*, and got his/her own unique trial or set of trials (as in Foulkes et al., Kanthamani & Broughton). Also, the planned comparison in (d) allows us to determine whether studies that tested a single 'star subject' performed better than multiple-percipient studies. There have been no prior attempts to determine these performance differences, so we will examine them in the present paper. Regarding statistical testing, we set alpha at .05, where alpha is the planned level at which outcomes will be considered significant. There was no need to adjust alpha for multiple analyses within hypotheses and sub-hypotheses as tests of same required only single tests each time.

The advantages of REM monitoring and awakening dream participants during REM sleep, have been discussed above. All the MDL studies featured that protocol. Some non-MDL studies also featured it, although in one case (Roe, Sherwood, Luke, & Farrell, 2002), dream-target (i.e., video clip) "transmission" occurred multiple times merely to 'guarantee' a REM test period though no empirical evidence of REM was recorded. In another case (Hearne, 1981a), REM was recorded but the participant was not awakened. Nevertheless, we class both protocols as REM-monitoring because they incorporated REM measures so that participants may have benefited from it. Hearne (1981b) monitored REM in one of three sets of trials (8 trials per set). For our REM analysis, our database will feature all relevant studies, including Hearne's (1981b) REM-subset of data.

Finally, a Bayesian analysis will be conducted on our dataset. Bayesian analysis in parapsychology is a recent approach aimed at providing an alternative to classical 'frequentist' or null hypothesis significance testing (NHST), although NHST is the main thrust of this paper (see Bem, Utts, & Johnson, 2011; Rouder & Morey, 2011; Tressoldi, 2011; Utts, Norris, Suess, & Johnson, 2010; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). However, the benefit of Bayesian meta-analysis lies in the fact that it provides a measure of the probability of a phenomenon (H1) and its non-existence (H0) called *posterior odds*, which are the product of summary statistics (so-called Bayes fac-

tors, or probability ratios) and prior odds. More specifically, we will perform Bayesian Parameter Estimation, which emphasizes what is referred to as the *explicit posterior probability distribution* of the parameter values (e.g., the mean, *SD*, *ES*); that is, the conditional probability that is assigned after the data obtained from experiments are taken into account and the prior probability is updated. In this Bayesian approach to assessing null values, we set up a range of parameter values, including the null value, and use Bayesian inference to compute the relative credibility of each of these values. These parameter values are referred to as *prior distributions*, or 'priors', and must be specified in the definition of the model to be tested. We will conduct this test because Rouder, Morey and Province (2013) reassessed the Storm et al. (2010) meta-analysis of free-response studies by conducting a Bayesian analysis, albeit on their own modified database. Rouder et al. (2013) remained "skeptical of the existence of psi" (p. 245) in spite of professing "a degree of support" for it. In response, Storm, Tressoldi and Utts (2013), after revising the database, conducted their own Bayesian analysis and showed that the psi effect, measured as a percentage hit-rate, lies somewhere between 26% and 32%, where MCE = 25%.

Given all of the above, the following planned hypotheses are proposed:

*Hypothesis 1:* Dream-ESP studies produce statistical evidence of a communications anomaly known as extrasensory perception (ESP) as measured by effect size *ES*.

*Hypothesis 2:* The Maimonides Dream Laboratory (MDL) studies differ from non-Maimonides (non-MDL) studies.

*Hypothesis 3:* Mean *z* scores and mean Effect Size (*ES*) values for dream-ESP studies are different for telepathy, clairvoyance, and precognition conditions.

*Hypothesis 4:* Dream-ESP studies using REM monitoring produce a higher mean *z* score and a higher mean *ES* value than dream-ESP studies that do not use REM monitoring.

*Hypothesis 5:* Dynamic targets in dream-ESP studies produce a higher mean *z* score and higher mean *ES* value than static targets.

*Hypothesis 6:* Mean *z* scores and mean Effect Size (*ES*) values vary between: (a) same-perceiver dream-ESP studies and different-perceiver dream-ESP studies; (b) same-agent dream-ESP studies and different-agent dream-ESP studies; (c) single-perceiver dream-ESP studies and multiple-perceiver dream-ESP studies; and (d) single-subject dream-ESP studies ($N_p$ = 1, where $N_p$ is number of percipients) and multiple-perceiver dream-ESP studies ($N_p$ > 1).

*Hypothesis 7:* Number of choices (*k*) per trial is positively related to *z* score in dream-ESP studies.

*Hypothesis 8:* Bayesian analysis of dream-ESP studies yields statistical evidence of a communications anomaly known as extra-sensory perception (ESP).

All studies in our meta-analysis are marked by asterisks in the Reference section and are listed alphabetically in Table A1 (see Appendix A). While the inclusion of single-subject studies in meta-analyses are said to create effect-size artifacts and other problems, and are therefore usually excluded in meta-analyses, we decided not to break with the tradition started by Child (1985), and continued by Sherwood and Roe (2003, 2013) who included in their reviews single-subject studies (e.g., Child, Kanthamani, & Sweeney, 1977; Kanthamani, Khilji, & Rustomji-Kerns, 1988; McLaren & Sargent, 1982). We believe test results on Hypothesis 6(d) may settle the issue over whether single-subject studies pose a serious threat to the validity of meta-analysis.

## 2. Method

### 2.1. Study Retrieval

The period of analysis was from 1966 (when laboratory research into dream ESP began) to 2016. The following major English-language peer-reviewed journals and peer-reviewed publications were searched and accessed for studies: *Biological Psychiatry, Dreaming, European Journal of Parapsychology, Experimental Medicine & Surgery, International Journal of Dream Research, International Journal of Parapsychology, Journal of Consciousness Studies, Journal of Nervous and Mental Disease, Journal of Parapsychology, Journal of the American Society for Psychical Research, Journal of the Society for Psychical Research, NeuroQuantology, Perceptual and Motor Skills, the Proceedings of the Annual Convention of the Parapsychological Association (Proceedings),* and *Research in Parapsychology*.

To find appropriate research articles online, we conducted Internet searches through EBSCOhost of the relevant databases, as well as PsycINFO and PsycARTICLES, and other relevant databases (e.g., Informit, Lexscien, and Web of Science). The following keywords and subject headings were entered in the search (singly and in combination): *anomalous cognition, clairvoyance, dreaming, dream, ESP, dream ESP, extrasensory perception, paranormal, parapsychology, precognition, psi,* and *telepathy*. Most of our Internet searches yielded the same studies as those already identified from the journal search. We adopted the following criteria regarding dream-ESP study selection:

- Studies must be peer-reviewed and published (we include papers published in the *Proceedings* as the whole papers, not just the abstracts, are peer-reviewed. Two studies were excluded as they were not peer-reviewed; Luke, 2002; Roe, Sherwood, & Farrell, 2007);
- Studies must involve telepathy, clairvoyance, or precognition;
- Studies must provide sufficient information (e.g., number of trials and hits, or outcome statistics) for the authors to calculate *z* scores and apply appropriate statistical tests and calculate *ES* as $z/\sqrt{n}$ (Formula 5 in Appendix B). Following the convention in parapsychology, we have used the effect size (*ES*) measure *r* (see Clark-Carter, 1997, pp. 550-551, 558). The correlation coefficient *r* is one of the most commonly used effect size measures (Prentice & Miller, 1992).
- Study targets must be randomly selected from experimenter-compiled target sets.

### 2.2. Procedure

For each study, we noted the following ten factors: (1) MDL/non-MDL (i.e., whether the study was conducted at the Maimonides Dream Lab or not); (2) type of dream-ESP task (i.e., clairvoyance, telepathy, or precognition); (3) the criteria adopted for selecting participants; (4) number of participants; (5) type of target (static vs. dynamic); (6) number of trials; (7) number of hits; (8) number of alternatives in the tasks; (9) REM-monitoring; and (10) agent/perceiver relationship. For studies in which *z* scores were not given, we calculated *z* scores from reported inferential statistics or effect sizes (see Appendix B). In two cases, the *z* score was calculated from a Sandler's A value (studies #12a and #12b), and in another case, the *z* score was calculated from a *p* value (study #50).

For the various statistical analyses, studies were grouped into: (1) MDL or non-MDL; (2) type of ESP (clairvoyance, telepathy, or precognition); (3) REM or non-REM; (4) static or dynamic targets; (5) same perceiver (or not); (6) same agent (or not); (7) single-perceiver (or not); (8) single-case and multiple-percipient studies, and (9) size of *k* (number of choices). The 'comparison' hypotheses listed in the previous section are based on these divisions.

Once the database was compiled, all studies were rated for quality by two judges who were kept 'blind' during the judging phase. We followed the protocol set down by Bem, Palmer, and Broughton (2001, p. 209), whereby the judges saw only the Method sections from which all identifiers had been deleted, such as article titles, authors' hypotheses, and references to results of other experiments in the article. The following seven criteria, adapted from Milton (1997), were adopted:

- Appropriate randomization (using electronic apparatuses or random tables);
- Random target positioning during judgment (i.e., target was randomly placed in the presentation with decoys);
- Blind response transcription or impossibility to know the target in advance;
- Number of trials pre-planned;
- Sensory shielding from agent (sender) and receiver (perceiver);
- Target independently checked by a second judge;
- Experimenter(s) blind to target identity.

The two judges answered "Yes" or "No" to each of the criteria. A study's mean quality score is the ratio of points awarded with respect to the items applicable (minimum score is 0/7 = 0.00; maximum score is 7/7 = 1.00). We stress that failure of studies to make explicit declarations of criteria does not mean that any given criterion was not incorporated into that experiment, but our approach is conservative so if there was no evidence in print that a criterion was met, the study was given a reduced quality rating.

Computer data were first entered into an MS Excel file, and later converted to an SPSS (Version 23) datafile for statistical analysis. Raw data were analysed statistically using SPSS programs and the online *VassarStats Binomial Probabilities* (Lowry, 2001-2014; for details, see Appendix B). Specific statistical tests for all hypotheses include One-Way Analysis of Variance (ANOVA) tests, Pearson's *r*, Spearman's *rho*, One-Sample *t* test, and Independent-Samples *t* test. These tests were planned in advance, as recorded in email correspondence between the authors (copies of relevant emails can be made available).

## 3. Results

### 3.1. Descriptive Statistics and Quality Ratings

We found 40 dream-ESP studies (totalling 52 datasets) conducted by 51 experimenters (see References for articles marked with asterisks that indicate inclusion in the meta-analysis; for dataset details, see Table A1 in Appendix A). The total trials count was 1,968 with 734 hits. Where *k* (i.e., the number of choices) was stipulated in studies, 26 out of 44 studies (59%) used a two-choice design (i.e., $k = 2$), and of these 26, 15 (58%) were MDL studies, targets of which were judged using the 'binary' (top-half/bottom-half) judging system described above. For the '$k = 2$' studies ($n = 24$), there were 787 trials and 438 hits, corresponding to a 56% hit rate, where mean chance expectation (MCE) = 50% ($z = 3.44$, $p < .001$). Where stipulated in studies, 14 out of 44 studies (32%) used a four-choice design (i.e., $k = 4$). For the '$k = 4$' studies ($n = 13$, since one study did not give a count of actual hits [study #16a in Appendix A, Table A1]), there were 867 eligible trials and 247 hits, corresponding to a 28% hit rate, where MCE = 25% ($z = 2.33$, $p = .01$).

Regarding quality ratings, Cronbach's alpha for the two judges' ratings was .84, suggesting a high degree of inter-rater reliability. When we averaged the two judges' quality ratings, case-by-case, and then calculated a mean value, we arrived at a score of 0.64 ($SD = 0.21$)—that is, on average, each study could be said to have met 4 or 5 criteria out of 7. The correlation between mean quality scores and *ES* values was extremely weak and not significant, $r(50) = .09$, $p = .527$ (two-tailed). As the database was later shown to be heterogeneous, we include here the equivalent nonparametric (Spearman's rho) statistics, $r_s(50) = .04$, $p = .770$ (two-tailed). Both results suggest *ES* is not likely to be an artifact of poor experimental design

Only 3 studies (6%) received a perfect score from at least one judge. Most criteria (at least 5), from at least one judge, were met in 36 of 52 studies (i.e., 69%). However, when we looked at combined judges' scores, most criteria (at least 5) were met in only 28 of 52 studies (i.e., 54%). Our conclusion is that these dream-ESP studies were not conducted with the same due care as the Ganzfeld studies which were quality-rated by the same judges (Storm et al., 2010), for which "17 studies (25%) received a perfect score from at least one judge, [and] most criteria (i.e., five or more out of seven) were met in 58 of 67 studies (i.e., 87%)" (p. 474).

Compared to the MDL dataset, the non-MDL dataset was superior in quality rating on all of three measures ('*perfect score from at least one judge*', '*at least 5 criteria met from combined judges*', and '*at least 5 criteria met from at least one judge*'): (i) Two studies from the non-MDL dataset received a perfect score from at least one judge compared to only one study from the MDL dataset; and (ii) from the combined judges' scores, most criteria (at least 5) were met in 21 of 37 studies (i.e., 57%) for the non-MDL dataset compared to 7 of 15 studies (i.e., 47%) for the MDL dataset, and (iii) most criteria (at least 5) from at least one judge, were met in 26 of 37 studies (i.e., 70%) for the non-MDL dataset compared to 10 of 15 studies (i.e., 67%) for the MDL dataset.

### 3.2. Planned Analyses

*H1: Dream-ESP studies produce statistical evidence of a communications anomaly known as ESP as measured by effect size ES.*

For the 52 studies that were drawn from 40 articles, we calculated a mean $z = 0.70$, a mean $ES = 0.18$, and Stouffer $Z = 5.01$, $p = 2.72 \times 10^{-7}$. Twelve out of 52 studies (23%) have positive $z$ scores that are independently significant. We note that only two studies (4%) have negative $z$ scores that are independently significant, bringing the total number of independently significant studies to 14 (27%). We did not include Watt, Wiseman, and Vuillaume (2015) in our database for the following reasons: First, in that study, there is an extra stimulus in the form of tactile material ("Each clip was linked to an object"; p. 175), thus forming a dual-target which is not categorical (i.e., not 'static' or 'dynamic')—one of the stimuli (the object) is static, and the other (video) is dynamic. Second, the judge is given a photo of the object in addition to the mentation, and this photo can be considered a psychometric aid to the judge's psi—in the standard Dream-ESP study, judges are asked to judge the mentation only. Thus, we cannot attribute ESP to the participant only. Hence, we regarded the study as a methodological outlier, and excluded the study from our analysis. However, if that study is included (6 hits our of 20 trials), changes to the outcome are marginal: $z$ score = 0.69, $ES = 0.18$, Stouffer $Z = 5.00$, $p = 2.87 \times 10^{-7}$.

The skew of the $z$-score distribution is significantly heterogeneous, although the skew of the distribution of $ES$ values was homogeneous. To test the skewness and kurtosis, the skewness and kurtosis values were divided by their respective $SE$ values, and if the statistics were between -1.96 and +1.96 (not significant), they were regarded as normal (George & Mallery, 2010). Outliers were identified from boxplots. In accordance with our conservative approach, two outliers were excluded from further analyses in the present article for having extreme $z$ scores—these are study #2a ($z$ = -6.08) and study #37b ($z = 4.64$) (see Appendix A, Table A1).

After the removal of the two outlier studies, a homogeneous dataset of 50 dream-ESP studies yielded mean $z$ = 0.75 ($SD = 1.12$; range: 4.82, min. = -1.64, max. = 3.18), mean $ES = 0.20$ ($SD = 0.31$; range: 1.34, min. = -0.40, max. = 0.94), and Stouffer $Z = 5.32$, $p = 5.19 \times 10^{-8}$. A One-Sample $t$ test (test value = 0.00) on $ES$ values was significant, $t(49) = 4.48$, $p < .001$ (two-tailed). Ninety-five percent confidence intervals (CIs) are as follows: $z$ scores, [0.43, 1.07]; $ES$ values, [0.11, 0.29]. Alternatively, we note that a random-effects analysis weighting $ES$ values for the inverse of their variance (which depend on $N$) gives $ES = 0.13$, 95% CI [0.08, 0.18]; $z = 4.9$, $p = 9.5 \times 10^{-7}$. Note that neither of these 95% CIs includes values of MCE (i.e., zero). Of the 50 studies, 39 (78%) had positive $z$ scores. Eleven (28%) of the 39 studies with positive $z$ scores (or 22% of the total set of 50 studies) are independently significant ($\alpha \leq .05$).

As Storm et al. (2010) advised, a more stringent measure than the Stouffer $Z$ for testing the overall significance of a database is provided by Darlington and Hayes (2000), who regard "mean($z$) as the real test statistic" (p. 505). We applied Darlington and Hayes's (2000) so-called 'Stouffer-max test' that provides a so-called MeanZ($s$, $k$) value which is the "mean of the s highest of $k$ mutually independent values of $z$" (p. 506). The outcome MeanZ is then compared with a critical MeanZ. Our MeanZ is 2.25. We took $s = 10$ (i.e., the ten studies with significant $z$ scores) and $k = 50$ (i.e., where $k = N = 50$). Darlington and Hayes (p. 506, Table 3) gives critical MeanZ = 1.73. In other words, the mean $z$ for the dream-ESP database is sufficiently higher than is required by the Stouffer-max test.

We can be conservative in our estimates using Darlington and Hayes's (2000, p. 503, Table 2) tabled data. If 10 individual results are significant with $\alpha = .05$, then pooled $p \leq .05$ only if the total number of studies is up to 110. In other words, we find a "fail-safe $N$" of 110 unpublished studies must exist in total in the file-drawer. Such a number of unpublished studies is unlikely to exist. This method does not require the assumption that the mean effect size of the missing studies is zero; "all the missing effect sizes may be highly negative" (p. 500).

*H2: The Maimonides Dream Laboratory (MDL) studies differ from non-Maimonides (non-MDL) studies.*

We tested this hypothesis by comparing respective mean $z$ scores, and mean $ES$ values. First, we divided the above homogeneous database ($N = 50$) into two groups—MDL and non-MDL datasets. Two Independent-Samples $t$ tests were conducted with mean $z$ scores and mean $ES$ values as the respective test variables, and 'dream laboratory' as the grouping variable.

There were 14 MDL studies and 36 non-MDL studies suitable for analysis. The MDL studies yielded mean $z = 0.90$, and mean $ES = 0.33$ ($SD = 0.37$), Stouffer $Z = 3.37$, $p = 3.76 \times 10^{-4}$. The non-MDL studies yielded mean $z = 0.69$, and mean $ES = 0.14$ ($SD = 0.27$), Stouffer $Z = 4.16$, $p = 1.57 \times 10^{-5}$. The $t$ test showed a difference in mean $ES$ values that was not significant, $t(48) = 1.97$, $p = .055$ (two-tailed), but the same test on mean $z$ scores was not significant, $t(48) = 0.58$, $p = .562$ (two-tailed). Hypothesis 2 was not supported.

*H3: Mean z scores and mean ES values for dream-ESP studies are different for telepathy, clairvoyance, and precognition conditions.*

Two One-Way Analysis of Variance tests were conducted with mean $z$ scores and mean $ES$ values as respective dependent variables, and 'type of ESP' as a fixed factor.

Table 1 lists by ESP modality the following statistics: mean $z$ scores, mean $ES$ values, Stouffer $Z$, and corresponding $p$ values. Across the three ESP categories, 25 studies (50%) tested telepathy, 13 studies (26%) tested clairvoyance, and 10 studies (20%) tested precognition. Studies 24 and 25 (Markwick & Beloff, 1983, 1988) were removed from the analysis because clairvoyance and precognition trials were combined. The ANOVA test on mean $z$ scores was not significant, $F(2, 47) = 0.17$, $p = .841$ (two-tailed). The ANOVA test on mean $ES$ values was also not significant, $F(2, 47) = 0.12$, $p = .889$ (two-tailed). Hypothesis 3 was not supported.

*H4: Dream-ESP studies using REM monitoring produce a higher mean z score and a higher mean ES value than dream-ESP studies that do not use REM monitoring.*

Two Independent-Samples $t$ tests were conducted with mean $z$ scores and mean $ES$ values as test variables, and REM as the grouping variable. Twenty-two studies tested ESP with REM monitoring, and 28 studies tested ESP without REM monitoring.

The 'REM-monitoring' studies yielded a mean $z$ score = 0.64 ($SD = 0.92$). The 'no-REM-monitoring' studies yielded a mean $z$ score = 0.84 ($SD = 1.26$). The difference be-

*Table 1*. Regression Models for the DIS Dream Quantity Score.

| | *Z* | | | Effect Size (*ES*) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **ESP Modality[b]** | ***M*** | ***SD*** | ***Skew*** | ***SE*** | ***M*** | ***SD*** | ***Skew*** | ***SE*** | **Sums of *Z* (∑*Z*)** | **Stouffer *Z*** | ***p*[a]** |
| Telepathy (*N* = 25) | 0.75 | 1.00 | 0.95 | 0.46 | 0.22 | 0.31 | 0.97 | 0.46 | 18.68 | 3.74 | $9.20 \times 10^{-5}$ |
| Clairvoyance (*N* = 13) | 0.88 | 1.39 | -0.23 | 0.62 | 0.18 | 0.33 | -0.27 | 0.62 | 11.46 | 3.18 | $7.36 \times 10^{-4}$ |
| Precognition (*N* = 10) | 0.59 | 1.20 | -0.39 | 0.69 | 0.17 | 0.34 | -0.39 | 0.69 | 5.94 | 1.88 | $3.00 \times 10^{-2}$ |

Note. [a]One-tailed; [b]Data from Markwick and Beloff (1988) are excluded because CL and PR trials were combined.

tween the two conditions was not significant, $t(48) = 0.62$, $p = .271$ (one-tailed).

The 'REM-monitoring' studies yielded mean *ES* = 0.24 (*SD* = 0.33). The 'no-REM-monitoring' studies yielded mean *ES* = 0.16 (*SD* = 0.29). The difference between the two conditions was not significant, $t(48) = 0.92$, $p = .180$ (one-tailed). Hypothesis 4 was not supported.

*H5: Dynamic targets in dream-ESP studies produce a higher mean z score and higher mean ES value than static targets.*

Two Independent-Samples *t* tests were conducted with mean *z* scores and mean *ES* values as test variables, and 'target type' as the grouping variable. There were 30 studies that used static targets, and 20 studies that used dynamic targets.

The dynamic-target studies yielded mean *z* = 1.00 (*SD* = 1.09). The static-target studies yielded mean *z* = 0.59 (*SD* = 1.12). Although in the direction hypothesized, the difference was not significant, $t(48) = 1.30$, $p = .100$ (one-tailed).

The dynamic-target studies yielded mean *ES* = 0.28 (*SD* = 0.33). The static-target studies yielded mean *ES* = 0.14 (*SD* = 0.29). Although the mean *ES* for studies with dynamic targets is twice that of the static-target studies, the difference was not significant, $t(48) = 1.51$, $p = .068$ (one-tailed). Hypothesis 5 was not supported.

*H6: Mean z scores and mean ES values vary between: (a) same-perceiver studies and different-perceiver studies; (b) same-agent studies and different-agent studies; (c) single-perceiver studies and multiple-perceiver studies; and (d) single-subject studies and multiple-perceiver studies.*

For (a), a One-Way ANOVA was conducted as there were three groups: (i) same-perceivers, (ii) different-perceivers, and (iii) a mix of both. Mean *z* score and mean *ES* were the dependent variables. For (b), (c), and (d), Independent-Samples *t* tests were conducted with mean *z* score and mean ES as the dependent variables, and the three dichotomous fixed factors were: (b) 'same-agent vs. different-agents, (c) 'single-perceiver vs. multiple-perceiver', and (d) '$N_p$ = 1' studies vs. '$N_p$ > 1' studies (where $N_p$ = number of perceivers).

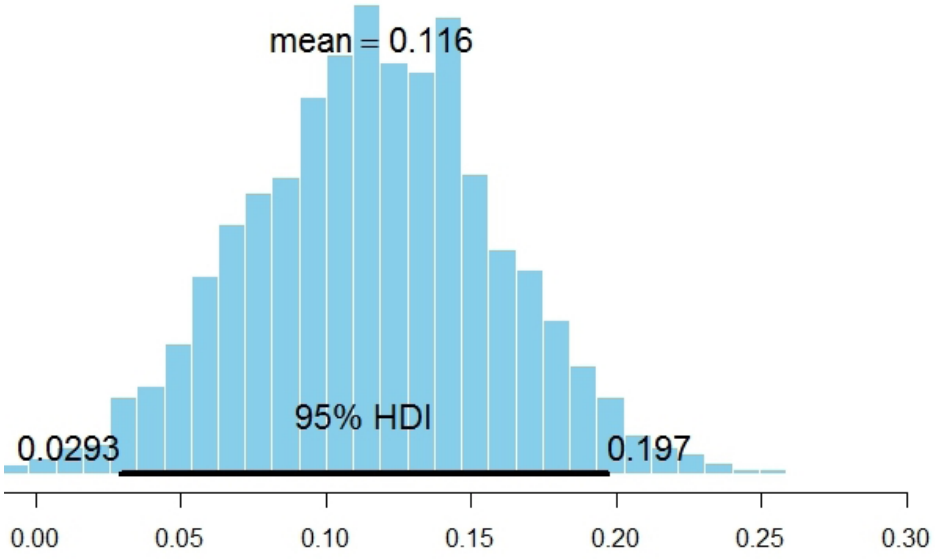
- (a) There were three conditions for this analysis: (i) 25 studies used the same perceiver; (ii) 13 studies did not use the same perceiver; and (iii) 12 studies used a mixed-perceiver condition. All 50 studies were in the analysis. The same-perceiver studies yielded mean *z* score = 0.69 (*SD* = 1.12), and *ES* = 0.20 (*SD* = 0.34). The different-perceiver studies yielded mean *z* score = 0.68 (*SD* = 1.21), and mean *ES* = 0.19 (*SD* = 0.32). The mixed-perceiver studies yielded mean *z* score = 0.96 (*SD* = 1.12), and mean *ES* = 0.20 (*SD* = 0.24). The mean *z* score difference between conditions was not significant, $F(2, 49) = 0.28$, $p = .760$ (two-tailed). The mean *ES* difference between conditions was not significant, $F(2, 49) = 0.004$, $p = .996$ (two-tailed).
- (b) The agent database was reduced by 25 studies that did not provide agent information. There were 15 same-agent studies and 10 different-agent studies suitable for analysis (total of 25 studies). The same-agent studies yielded mean *z* score = 0.88 (*SD* = 1.06), and mean *ES* = 0.24 (*SD* = 0.30). The different-agent studies yielded mean *z* score = 0.55 (*SD* = 0.93), and mean *ES* = 0.20 (*SD* = 0.34). The mean *z* score difference between the two conditions was not significant, $t(23) = 0.78$, $p = .445$ (two-tailed). The mean *ES* difference between the two conditions was not significant, $t(23) = 0.32$, $p = .750$ (two-tailed).
- (c) There were 34 single-perceiver studies and 16 multiple-perceiver studies suitable for analysis (total of 50 studies). The single-perceiver studies yielded mean *z* score = 0.60 (*SD* = 1.12), and mean *ES* = 0.20 (*SD* = 0.35). The multiple-perceiver studies yielded mean z score = 1.07 (*SD* = 1.08), and mean *ES* = 0.18 (*SD* = 0.19). The mean *z* score difference between the two conditions was not significant, $t(48) = 1.37$, $p = .176$ (two-tailed). For the test on mean *ES* values, equal variances was not assumed, Levene's $F = 5.48$, $p = .023$. The difference between mean *ES* values between the two conditions was not significant, $t(46.85) = 0.20$, $p = .840$ (two-tailed).
- (d) There were 20 '$N_p$ = 1' studies and 30 '$N_p$ > 1' studies suitable for analysis (total of 50 studies). The '$N_p$ = 1' studies yielded mean *z* score = 0.58 (*SD* = 1.12), and mean *ES* = 0.19 (*SD* = 0.37). The '$N_p$ > 1' studies yielded mean *z* score = 0.87 (*SD* = 1.12), and mean *ES* = 0.20 (*SD* = 0.27). The mean *z* score difference between the two conditions was not significant, $t(48) = 0.91$, $p = .369$ (two-tailed). For the test on mean *ES* values, equal variances was not assumed, Levene's $F = 4.09$, $p = .049$. The difference between the two conditions was not significant, $t(32.09) = 0.06$, $p = .951$ (two-tailed).

The four-part hypothesis was not supported.

*H7. Number of choices (k) per trial is positively related to z score in dream-ESP studies.*

There were three sub-sets of studies: '*k* = 2' (*n* = 24; Mean *z* score = 0.79), '*k* = 4' (*n* = 12; Mean *z* score = 0.52), and '*k* = 5' (*n* = 2; Mean *z* score = 0.76). Only one study had a

*Figure 2*. Highest Density Interval related to *ES*. All values inside the interval, indicated by the heavy black horizontal line, have higher credibility than values outside the interval, where the interval includes 95% of the respective distribution.




'*k* = 3' design so that study was excluded (Study #28); and one study used a '*k* = 4445' design (Study #8; a number was chosen between 1111 and 5555; therefore, *k* = [5555 – 1111] +1 = 4555), which was also excluded. In accordance with Timm's (2000) conjecture, increasing the number of target alternatives *k* must lead to increased levels of significance in ESP experiments, which can be discerned from *z* scores. However, the mean *z* scores indicate a U shape and therefore do no suggest an incline as *k* increases. A Jonckheere-Terpstra test for ordered alternatives shows no significant trend of higher *z* scores as *k* increases, $T_{JT}$ = 192.50, *z* = -0.37, *p* = .713 (two-tailed). A non-significant Kendall's tau-b also shows a very weak negative effect, $r_{\tau}(38)$ = -.05, *p* = .713 (two-tailed). There was no evidence that the number of choices affects *z*-score outcome.

*H8: Bayesian analysis of dream-ESP studies yields statistical evidence of dream ESP.*

We used Bayesian parameter estimation to analyze the 50 studies in our database in order to test whether our significant finding yielded by a classical frequentist (NHST) approach and random effects model was upheld (see the test results for *H1* above). In Bayesian analyses, θ is conceptualized as a random variable and, thus, is considered to fluctuate over time (Field, 2009). More specifically, θ is the parameter we are attempting to estimate based on our database. In the present study, our prior subjective belief regarding θ was 0.1. The prior distribution of this parameter, θ, was specified as normal. The measure of variance among the different studies formed a uniform distribution from –1 to +1, estimated with 50,000 Markov Chain Monte Carlo iterations, each starting with different and dispersed initial values for the model. We were not interested in the estimate of the *SD*.

Figure 2 depicts the posterior probability distribution of θ. That is, the conditional probability that is assigned after the *ES* of each of the 50 experiments is taken into account and the prior probability of θ is updated. The across-experiments μ (i.e., average *ES*; see Figure 2) was above chance. The 95% Highest Density Interval (HDI) of posterior probability (which indicates the most plausible 95% of the values in the posterior distribution) related to the *ES* ranged from 0.03 to 0.20 (mean *ES* = 0.12). The HDI indicates the most plausible 95% of the values in the posterior probability distribution. As shown in Figure 2, θ is included in the 95% HDI. Thus, the null hypothesis is rejected confirming the results obtained with the frequentist random model.

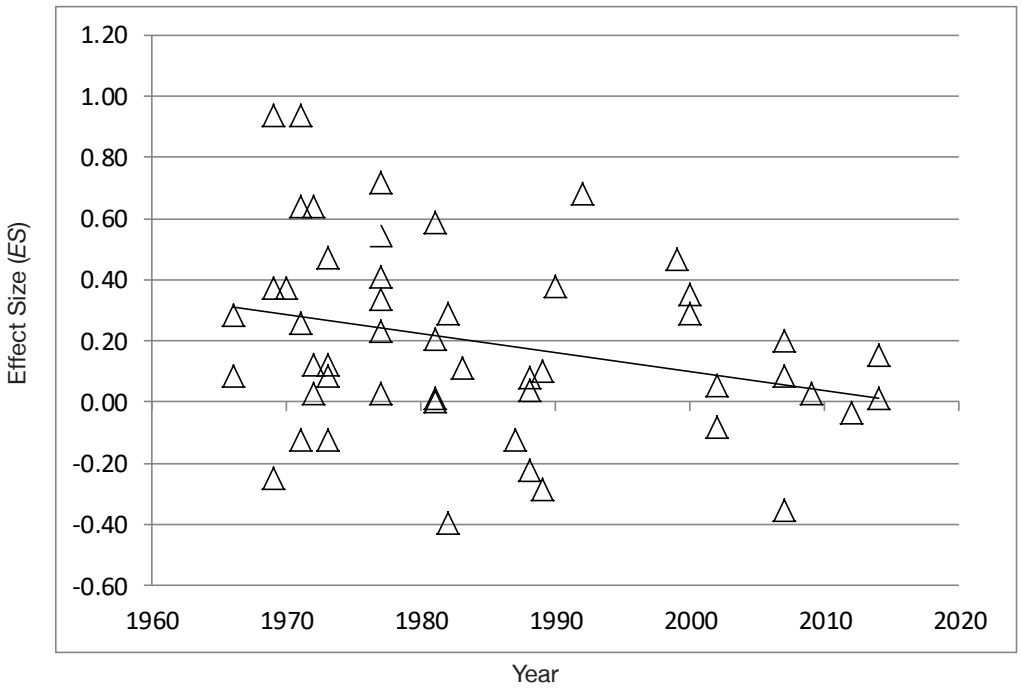
## 3.3. Post Hoc Findings

### *3.3.1 Experimenter/laboratory ES comparisons.*

According to Akers (1987), psi effects have not been replicated amongst the majority of investigators. To ascertain whether our database overall was the result of extremely positive *ES* values for 'pockets' of experimenters/laboratories, rather than a general trend across experimenters/labs, we conducted a One-Way ANOVA on the pooled data (*N* = 50) after dividing them into experimenter/laboratory groups. We could not test experimenter/laboratory interaction as we found that a number of experimenters had worked in more than one laboratory.

We formed 12 mutually exclusive experimenter/laboratory groups with at least two studies in each: "Belvedere," "Braud," "Child," "Dalton," "Harley," "Hearne," "Kanthamani", "Krippner", "Luke", "Markwick"," "Roe", and a twelfth "Miscellaneous" group comprised of left-over stud-

*Figure 3*. Scatterplot of *ES* values for dream-ESP studies over a period of 51 years (1966-2016). A significant decline ($r_s$ = -0.29) is indicated ($p$ = .044).




ies. Mean *ES* values ranged from zero to 0.63, but no group was significantly different, $F(11, 49) = 1.21$, $p = .314$ (two-tailed). The effects cannot be said to be due to a few outstanding investigators.

### 3.3.2 *Decline effects across time in the dream-ESP databases.*

As can be seen from the above review, procedures for the studies during this period were mixed and complex, involving a number of researchers with different methodologies, different statistical testing procedures, and different goals. A few researchers tested an ambiguous form of telepathy/clairvoyance known as *general extra-sensory perception* (GESP), while others tested precognitive dreaming or clairvoyant dreaming.

In spite of these differences, we decided to perform one final analysis involving the assessment of a hypothesized *ES* decline over the time period 1966 to 2014 (no dream ESP studies could be found for the period 2015 to 2016). In combination with the results for quality ratings given earlier, and the non-significant relationship between quality ratings and *ES* values, any evidence of a decline would help establish whether or not the dream-ESP research is as successful or as fruitful as the above results suggest.

Figure 3 shows *ES* values plotted for the 50 dream-ESP studies. We note that the correlation between year of study and *ES* is negative and significant, $r(48) = -0.29$, $p = .044$ (two-tailed). This result indicates a linear decline in *ES*s over the 49-year period, as illustrated in Figure 3. This linear decline is formulated thus: *ES* = [0.0062*YEAR] + 12.576, $R^2 = 0.08$. A re-test using our homogeneous data shows no significant decline in *ES* related to quality, $r(48) = .08$, $p = .600$ (two-tailed), but quality control in experiments has improved over the 49-year period, $r(48) = 0.39$, $p = .006$ (two-tailed). As an exercise, if we partial out the effect of year on quality and *ES*, we find a *positive* relationship between quality and *ES* approaching significant, $r(47) = .21$, $p = .073$ (one-tailed). This result is worthwhile recording here insofar as it illustrates the complete opposite of the sceptical hypothesis that improvements in quality necessarily mean *ES* must plummet (see Rao, 2001, for similar findings).

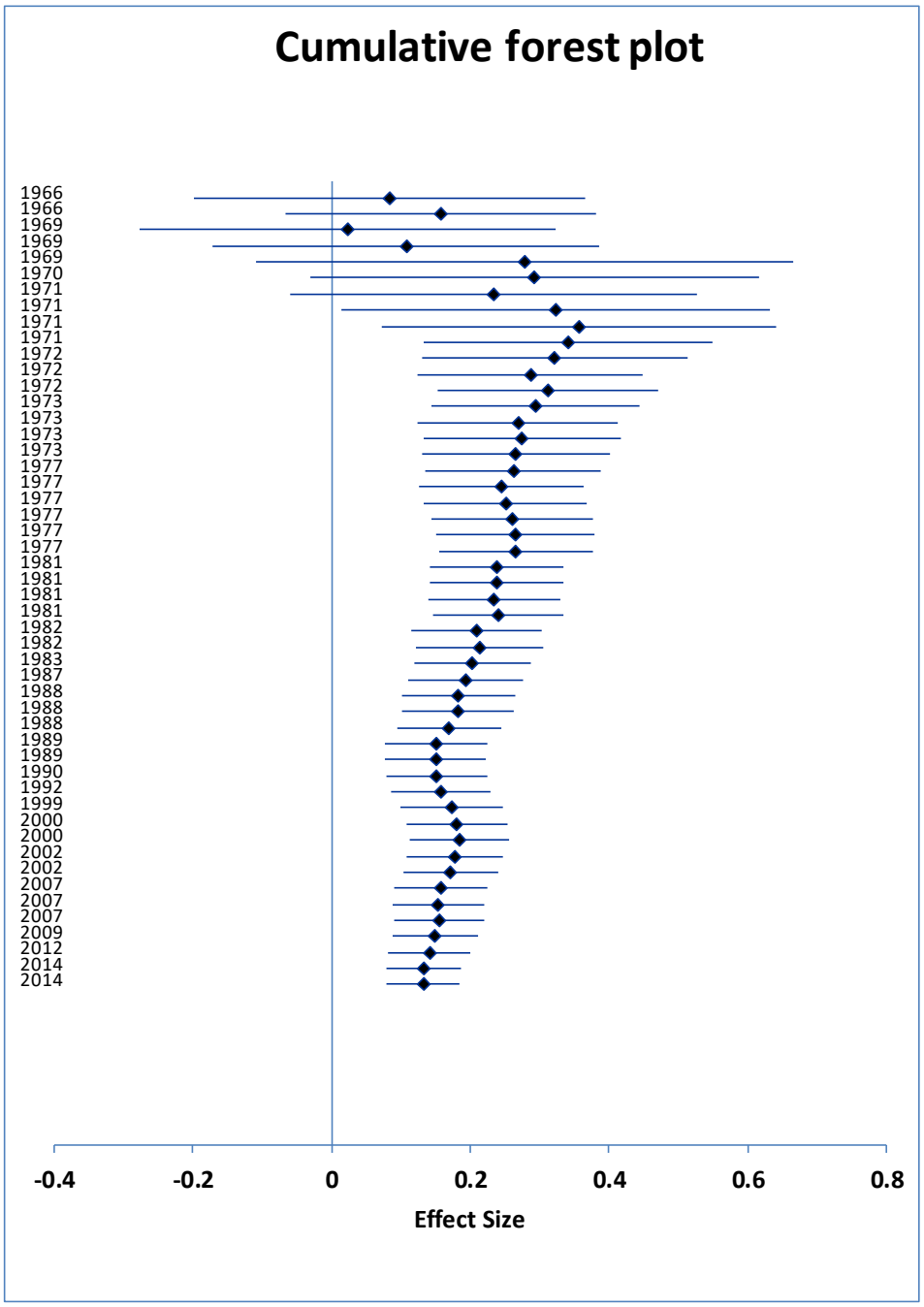
A forest plot, which is a cumulative representation of studies illustrating possible time trends in effect sizes as new studies are added, was generated to show shifts in the cumulative weight of the evidence over time (Rothstein, Sutton, & Borenstein, 2005). We added studies successively by their publication year and Figure 4 shows a tendency for *ES* values to increase and then go into decline, thus reflecting the above findings. Although Figure 4 suggests that the accumulation of effects over time may be attributable to older *ES* values, we emphasize that our other analysis (see the section *Descriptive Statistics and Quality Ratings*) does not allow us to attribute this decline to improved study quality.

## 4. Discussion

We have shown that our homogeneous 50-study dream-ESP database yields a mean *z* of 0.75, a mean *ES* of 0.20, and is significant overall, Stouffer $Z = 5.32$, $p = 5.19 \times 10^{-8}$ (see other test results for *H1* above). We note that an *ES* of 0.20 falls about midway in the range of *ES* values of five independent Ganzfeld meta-analyses (see Table 2). Our mean *z* score (0.75), however, is overall weaker than all five Ganzfeld mean *z* scores, and this discrepancy is attributable to sample-size differences between the two types of experiment.

Critics may regard these parapsychological effects as generally weak, but weak effects are often reported in the research literature, yet major decisions are made on the strength of the results. For example, a weak effect ($r = .03$) was reported in the aspirin/heart-attack study by the Steer-

*Figure 4*. Forest plot showing that, accumulatively, effect sizes have remained above chance on average, but are in decline from 1966 to 2014 (NB: no dream ESP studies were found for the period 2015 to 2016).




ing Committee of the Physicians' Health Study Research Group (1988), but the study was halted on ethical grounds because 45% fewer heart attacks were reported in the experimental group compared to the control group. Likewise, on ethical grounds, the National Heart, Lung, and Blood Institute (DeMets, Hardy, Friedman, & Lan, 1984; Kolata, 1981) discontinued a study nine months ahead of schedule because of the clear benefits of propranolol on patients with a recent myocardial infarction, even though the effects were also reported to be weak ($r$ = .04). These are but two examples (see also Spencer, 1995).

We reported an effect *ES* twice the size for the MDL studies (*ES* = 0.33) compared with the non-MDL studies (ES = 0.14), but the difference was not significant (see *H2*). However, the Stouffer *Z* values for both databases were significant. Of the Maimonides (MDL) studies, Child (1985) concluded that chance could not explain the results, and he argued that there was "some systematic—that is, nonrandom—source of anomalous resemblance of dreams to targets" (Child, 1985, p. 1222; and see again test results for *H2* above). Sherwood and Roe (2013) concluded rather broadly that "Combined effect sizes for both Maimonides and post-Maimonides studies suggest that judges may be able to use dream mentations to correctly identify target materials more often than would be expected by chance" (p. 44), and they singled out as exceptional the 'sensory bombardment' te-

*Table 2*. Performance Comparisons: Five Independent Ganzfeld Meta-Analyses

| Author (Year) | Studies | Hit % | *ES* | Mean *z* | *p* |
|---|---|---|---|---|---|
| 1. Bem & Honorton (1994) | 10 | 32.2 | 0.61 | 2.89 | .002 |
| 2. Bem, Palmer, & Broughton (2001) | 10 | 36.7 | 0.13 | 1.26 | .104 |
| 3. Honorton (1985)[a] | 28 | 35.0 | 0.24 | 1.25 | .107 |
| 4. Storm & Ertel (2001)[b] | 11 | 31.6 | 0.21 | 0.91 | .181 |
| 5. Storm, Tressoldi, & Di Risio (2010) | 29 | 32.2 | 0.14 | 1.02 | .154 |

Note. [a]Cited in Bem and Honorton (1994); [b]z scores and effect sizes are adjusted from those given in Storm and Ertel's Table 1 (2001, p. 428).

lepathy study by Krippner et al. (1971; study #19 in Table A1). We would add Ullman and Krippner (1969) as another outstanding study (study #35 in Table A1), which also happens to be a telepathy study. Some precognition studies were also very successful (e.g., studies #17, #21, & #31 in Table A1). One may also add the recent non-MDL study by Watt (2014) to that short-list of impressive precognition studies (NB: our opinion remains unchanged given the later inclusion of unreported missing data—see Watt & Valášek, 2015). Watt (2014) originally excluded a subset of inadequate data that did not meet her pre-specified criterion (i.e., "data from any participants who did not complete four trials were discarded"; p. 105). We add that the note does not meet two of our criteria: the note does not provide sufficient information, and it was not adequately peer-reviewed.

Given the results of the present meta-analysis, we concur with Child (1985), and Sherwood and Roe (2013), with only slight reservations: First, it was the clairvoyance studies that produced the highest mean *ES* score, with precognition yielding the lowest (telepathy scoring mid-way), although the difference was not significant (see Table 1 for other results). Also, all three modalities yielded a significant Stouffer *Z*. Thus (to pre-empt pertinent statements in the next paragraph), while the three ESP modalities produce gainful outcomes, we cannot recommend a specific ESP modality to researchers, although we would add that precognition studies, which do not need a sender, are easier to run, and resist real-time sensory leakage of information about target identity, may be less costly and time-consuming to operationalize for those reasons alone.

A number of sub-group comparisons yielded null results. First, there was the test for *ES* differences between telepathy, clairvoyance, and precognition sub-sets (see *H3* test results). As was proposed in the Introduction, past results of no modality differences (Steinkamp, Milton, & Morris, 1998; Storm et al., 2010, 2012), coupled with this new null finding, suggest that ESP is either a consistent effect across the three modalities, indicating a possible ESP limitation no matter what form it takes, or ESP is manifested in only one way, and the three modalities should be considered expressions of a single underlying psi phenomenon or function. This commonality is reflected in a number of theoretical attempts to subsume different forms of psi under a single rubric, including Jung's (1960) Synchronicity, Thalbourne's (2004) Psychopraxia, May et al.'s (1995) Decision Augmentation Theory, and possibly Stanford's (1978) Psi-Mediated Instrumental Response model.

We also tested REM-monitoring studies against no-REM-monitoring studies, and found no difference (see test results for *H4*). Put simply, while it is argued that REM guarantees dreaming, it does not guarantee heightened ESP. As we said above, the advantage of awakening participants from REM sleep is that dream recall is much more likely, with more detailed and longer overall reports. For the researcher, that translates as an incentive that trumps non-REM awakening. While holding the expected reservations warranted from statistical testing, we would argue that the mean *ES* is encouragingly high for studies with REM-monitoring and researchers may wish to take that finding into consideration.

The long history of Ganzfeld research suggests dynamic targets are superior to static targets, but we could not confirm that statistically when testing the dream-ESP database. We do, however, concede as worthy of note the effect-size difference, for which—though not significant ($p$ = .068)—the *ES* for the dynamic-target-set of studies (0.28) was twice as high as the *ES* for the static-target-set (0.14; see test results for *H5*). For those reasons, we are not yet ready to abandon the consensus view that dynamic targets make a difference in ESP research (see also, Parker, 2005, for his arguments).

We note that Sherwood and Roe (2013) were not certain that a sender (i.e., agent) is necessary in dream-ESP studies. Similarly, we did not find that using the same agent repeatedly, as opposed to changing agents constantly, or testing a number of different agents in a study, made any significant difference to the mean *z* scores or mean *ES* values. Perhaps greater effort should go towards improving clairvoyance and precognition designs that do not need agents, especially given the fact that pure telepathy is currently not testable under laboratory conditions where controls are necessary (see Bauer's, 1984, comments above). Also, we did not find that it mattered if the perceiver stayed the same or was changed.

Along similar lines, we found that single perceivers did not perform better than multiple perceivers working together (*H6[c]*), and single-case studies ($N_p$ = 1) built around claimed psychics with 'star' status did not have a better track record where *ES* values were concerned than '$N_p$ > 1' studies comprised mostly of randomly selected 'non-claimants' (see test results for *H6[d]*). This finding suggests there is little point wasting time and money in the search for, and/or training of star subjects—at least for dream-ESP studies.

Also, we did not find a significant difference between levels of *k* (number of choices; see test results for *H7*). This hypothesis may not have found support because the subsamples were too small, with a range of only three levels of *k*. Therefore, we do not dismiss Timm's (2000) claim just yet, especially since the effect has been demonstrated for the forced-choice domain (Storm et al., 2012).

### 4.1. The Bayesian Analysis

We used a Bayesian approach to estimate the probability of the mean *ES* (see test results for *H8*). The 95% HDI of posterior probabilities concerning the *ES* did not include zero. Thus, the null hypothesis was rejected. Importantly, these Bayesian findings support the significant result we obtained using a classical frequentist approach. Nevertheless, we are mindful that Utts, Norris, Suess and Johnson (2010) cautioned that, "Bayesian methods utilize [a] 'degree of belief' interpretation of probability to model all uncertainty" (p. 2). Indeed, the statistician Gelman (2008) contended that, "as scientists we should be concerned with objective knowledge rather than subjective belief " (p. 2).

### 4.2. Criticisms of the MDL and Non-MDL Studies

Alcock (1981) criticized the MDL studies for lacking control groups, but the controls in such studies are the other non-target stimuli against which the transcript is also compared. The same is true for the non-MDL studies. The potential for multiple analyses that would serve to inflate the potential for a family-wise Type I error has also been raised (Child, 1985; Parker, 1975, p. 89), but this is addressed in meta-analysis, where a pre-specified common outcome statistic is used. Fraud has also been suggested as a possible explanation for the results (e.g., Clemmer, 1986), but no plausible mechanism for fraud has been put forward. One other defense against the implication of fraud is the result of our test between independent experimenters/laboratories, which contradicts Akers's (1987) claim that psi is an effect not replicated amongst the majority of investigators (see the test result in the above section *Post Hoc Findings*).

Also worthy of note is the decline effect (see test result in *Post Hoc Findings*), which is anathema to some parapsychologists, with some researchers insisting that all psi effects inevitably go into decline (Bierman, 2001; Bierman, Bosga, Gerding, & Wezelman, 1993; Milton & Wiseman, 1999). Other researchers point out that effects in mainstream psychology and other disciplines, also go into decline (Ioannidis, 2005; Schooler, 2011). As Bierman (2001) has indicated, however, the Ganzfeld effect across time happens to indicate a so-called rebound effect (in the form of a significant U-shaped curve), suggesting the effect is on the increase after a temporary slump only. Storm et al. (2010) also showed a rebound effect for the Ganzfeld, and Storm et al. (2012) found an incline in their forced-choice meta-analysis. Perhaps, it remains to be seen what direction the dream-ESP database will take over a much longer term than tested in this study, but although we found evidence of a decline in *ES* across time, there was no evidence that the decline in *ES* is attributable to improvements in design quality. In other words, as we said above, *ES* is not likely to be an artifact of poor experimental design. Figures 3 and 4 suggest that a large and consistent number of high-*ES* studies will be needed to cancel out the decline.

### 4.3. Conclusion and Recommendations

Our review has shown that dream ESP remains a promising, if somewhat neglected, area for parapsychological research. Combined effect sizes for both Maimonides and post-Maimonides studies suggest that judges may be able to use dream mentations to identify target materials correctly more often than would be expected by chance. There is evidence of conceptual replication within both sets of studies, and the effects seem not to be concentrated within certain research teams.

Sherwood and Roe (2013) concluded that the Maimonides studies were more successful than the post-Maimonides studies, and attributed that difference to "procedural differences rather than improvements in security" (p. 72). This may not be entirely true. Our results do not support claims of MDL success over non-MDL studies, though we do concede that other test findings *suggest* the MDL series may have been superior. But we must take serious stock in the quality issues raised above which go against the MDL studies (see especially the last paragraph in the section *Descriptive Statistics and Quality Ratings*). While both datasets, MDL and non-MDL, produced independently significant Stouffer *Z* values, others may agree with Sherwood and Roe and dispute our conclusion of no difference (see the test results for *H2*), and it could be argued that a larger database with no other changes, or adjusting $\alpha$ to $\leq .10$ because *N* is 'small', would see a significant difference. Nevertheless, the researcher has to decide whether the MDL differences will make a difference, and it is questionable whether all MDL-related conditions, from many decades ago, using specific personnel, can even be replicated. We do agree that "procedural differences", such as using dynamic targets (as suggested by the results for *H5*), may be advantageous to the ESP effect, and testing during REM-sleep may guarantee dreaming, which is an important consideration. And we have shown that even though there has been an *ES* decline over nearly five decades of dream-ESP research, the decline is not related to improvements in quality thus ruling out a potential artifact of poor design.

Our meta-analysis has identified key issues and key concerns to do mainly with methodological quality and process-oriented factors that covary with study outcomes. However, the database may prove to be too heterogeneous, sometimes with too few studies in subsets, for such analyses to provide reliable insights. We hope that future researchers will note some of the methodological shortcomings we have identified and address these in their study designs. We also hope that this review will help re-awaken interest in this neglected but promising paradigm, and we make the following additional recommendations:

1. There is a need for more systematic research programs in this area, involving confirmatory as well as exploratory and pilot studies;
2. It would be useful to investigate the efficacy of deliberate vs. natural awakening (and the use of home 'dream machines'), different judging techniques (e.g., participant vs. blind judges, individual vs. consensus), the effects of differing amounts of information (e.g., dreams only vs. dreams plus additional material), the efficacy of emotional vs. non-emotional targets;
3. Participants should arguably have good dream recall and should be allowed a pre-specified number of pilot trials in order to facilitate adjustment to the experimental procedure;
4. Details of experiments should be published in full (see Milton & Wiseman, 1997) in journals, and not just in parapsychology journals (to allow for wider dissemination and encourage larger numbers of replication or confirmation attempts);
5. Hypotheses and planned analyses should be clearly

stated and researchers should avoid multiple analyses using different outcome measures;
6. Full details of the procedure and the results of statistical analyses should be included in such reports. If target rankings are available, these should be summarized, even if direct or binary hit analysis is planned, so that they are available for future meta-analyses.
7. If a study is intended to be confirmatory, it should be pre-registered.

In closing, although our database may have its imperfections, and alternative analyses are possible, we made the standard meta-analytic modifications at every step. Our database thus far is comprehensive and, it should be mentioned, we took a conservative approach throughout, which allows us to make some justifiable, if not reliable, conclusions. We have seen that our findings made it necessary to update or modify some earlier conclusions and taken-for-granted assumptions. In particular, we would now say that dream ESP is (i) a demonstrable effect; (ii) not governed by experimenter, or laboratory, or historical context; (iii) independent of (a) psi modality; (b) REM monitoring; (c) target type; and (d) agent and perceiver arrangements; and (iv) perhaps independent of the number of choices in a target set. Some of these findings conflict with what we find to be evident of the free-response paradigm (including Ganzfeld) and the forced-choice paradigm, and it remains to be seen if our conclusions are premature, or dream ESP is, in a number of ways, an ESP sub-type different in degree or kind.

## References


Alcock, J. E. (1981). Parapsychology: science or magic? A psychological perspective. Elmsford, NY: Pergamon Press.

Akers, C. (1987). Parapsychology is science, but its findings are inconclusive. Behavioral and Brain Sciences, 10, 566-568.

Bauer, E (1984). Criticism and controversy in parapsychology: An overview. European Journal of Parapsychology, 5, 141-166.

* Belvedere, E., & Foulkes, D. (1971). Telepathy and dreams: A failure to replicate. Perceptual and Motor Skills, 33, 783-789.

Bem, D. J. (1993). The ganzfeld experiment. Journal of Parapsychology, 57, 101-110.

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

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

Bem, D. J., Utts, J., & Johnson, W. O. (2011). Must psychologists change the way they analyze their data? Journal of Personality and Social Psychology, 101, 716-719. doi:10.1037/a0024777

Bierman, D. J. (2001). On the nature of anomalous phenomena: Another reality between the world of subjective consciousness and the objective world of physics? In P. Van Loocke (Ed.), The physical nature of consciousness (pp. 269–292). New York, NY: Benjamins.

Bierman, D. J., Bosga, D., Gerding, J. L. F., & Wezelman, R. (1993). Anomalous information access in the Ganzfeld: Utrecht—Novice series I and II. In Proceedings of the 36th Annual Convention of the Parapsychological Association (pp. 192-204). Durham, NC: Parapsychological Association.

* Braud, W. (1977). Long-distance dream and presleep telepathy. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), Research in parapsychology 1976 (pp. 154-155). Metuchen, NJ: Scarecrow Press.

Braud, W.G., & Braud, L.W. (1975). The psi-conducive syndrome: Free response GESP performance following evocation of 'left-hemispheric' vs. 'right-hemispheric' functioning.' In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), Research in parapsychology 1974 (pp. 17-20). Metuchen, NJ: Scarecrow Press.

Child, I. L. (1985). Psychology and anomalous observations: The question of ESP in dreams. American Psychologist, 40, 1219-1230.

* Child, I. L., Kanthamani, H., & Sweeney, V. M. (1977). A simplified experiment in dream telepathy. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.), Research in parapsychology 1976 (pp. 91-93). Metuchen, NJ: Scarecrow Press.

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

Clemmer, E. J. (1986). Not so anomalous observations question ESP in dreams. American Psychologist, 41, 1173-1174.

* Dalton, K., Steinkamp, F., & Sherwood, S. J. (1999). A dream GESP experiment using dynamic targets and consensus vote. Journal of the American Society for Psychical Research, 93, 145-166.

* Dalton, K., Utts, J., Novotny, G., Sickafoose, L., Burrone, J., & Phillips, C. (2000). Dream GESP and consensus vote: A replication. Proceedings of the 43rd Annual Convention of the Parapsychological Association (pp. 74-85), Freiburg, Germany.

Darlington, R. B., & Hayes, A. F. (2000). Combining independent p values: Extensions of the Stouffer and binomial methods. Psychological Methods, 5(4), 496-515.

Dement, W. C. (1974). Some must watch while some must sleep. San Francisco, CA: W.H. Freeman.

Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. Electroencephalography and Clinical Neurophysiology, 9, 673-690.

DeMets, D. L., Hardy, R., Friedman, L. M., & Lan, K. K. G. (1984). Statistical aspects of early termination in the Beta-Blocker heart attack trial. Control Clinical Trials, 5, 362-372.

Empson, J. (2002). Sleep & dreaming (3rd ed.). New York, NY: Palgrave.

Eppinger, R. (2001). A comparative examination of ganzfeld and dream reports in free response ESP studies Unpublished PhD thesis, University of Edinburgh.

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

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). Los Angeles: Sage.

* Foulkes, D., Belvedere, E., Masters, R. E. L., Houston, J., Krippner, S., Honorton, C., & Ullman, M. (1972). Long-distance 'sensory-bombardment' ESP in dreams: A failure to replicate. Perceptual and Motor Skills, 35, 731-734.

George, D., & Mallery, M. (2010). SPSS for Windows step by step: A simple guide and reference. Boston, MA: Pearson.

Gelman, A. (2008). Objections to Bayesian statistics. Retrieved, March 13, 2011, from: ba.stat.cmu.edu/journal/2008/vol03/issue03/gelman.pdf.

Globus, G., Knapp, P. H., Skinner, J. C., & Healey, G. (1968). An appraisal of telepathic communication in dreams. Psychophysiology, 4, 365.

Goodenough, D. R. (1991). Dream recall: History and current status of the field. In S. J. Ellman, & J. S. Antrobus (Eds.), The mind in sleep: Psychology and psychophysiology (2nd ed.; pp.143-171). New York: John Wiley.

Hall, C. E. (1967). Experimente zur telepathischen Beeinflussung von Träumen [Experiments on telepathically influenced dreams]. Zeitschrift für Parapsychology und Grenzgebeite der Psychologie [Journal for Parapsychology and Frontier Areas of Psychology], 10, 18-47.

* Harley, T. A. (1989). Psi missing in a dream clairvoyance experiment. Journal of the Society for Psychical Research, 56, 1-7.

* Hearne, K. M. T. (1981a). Lucid dreams and ESP: An initial experiment using one subject. Journal of the Society for Psychical Research, 51, 7-11.

* Hearne, K. M. T. (1981b). The effect on the subject (in waking, SWS and REM states) of electric shocks to the agent: An 'ESP' experiment. Journal of the Society for Psychical Research, 51, 87-92.

* Hearne, K. M. T. (1987). A dream-telepathy experiment using a home 'dream machine. Journal of the Society for Psychical Research, 54, 139-142.

* Hearne, K. M. T. (1989). A nationwide mass dream-telepathy experiment. Journal of the Society for Psychical Research, 55, 271-274.

* Hearne, K. M. T., & Worsley, A. (1977). An experiment in 'telepathic' phobic fear and REM sleep. Journal of the Society for Psychical Research, 49, 434-439.

Honorton, C. (1977). Psi and internal attention states. In B.B. Wolman (Ed.), Handbook of parapsychology (pp. 435-472). New York, NY: Van Nostrand Reinhold.

Honorton, C., Berger, R., Varvoglis, M., Quant, M., Derr, P., Schechter, E., & Ferrari, D. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. Journal of Parapsychology, 54, 99-139.

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

* Honorton, C., Krippner, S., & Ullman, M. (1972). Telepathic perception of art prints under two conditions. Proceedings of the 80th Annual Convention of the American Psychological Association, 7, 319-320.

Honorton, C., Ullman, M., & Krippner, S. (1975). Comparison of extrasensory and presleep influences on dreams: A preliminary report. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), Research in parapsychology 1974 (pp. 82-84). Metuchen, NJ: Scarecrow Press.

Ioannidis, J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. JAMA: The Journal of the American Medical Association, 294(2), 218-228.

Ivtzan, I. (2007). Tarot cards: A literature review and evaluation of psychic versus psychological explanations. Journal of Parapsychology, 71, 139-149.

Ivtzan, I., & French, C. C. (2004). Testing the validity of tarot cards: Can we distinguish between a real and a control reading? Proceedings of the 47th Annual Convention of the Parapsychological Association (pp. 5-8). Durham, NC: Parapsychological Association.

Jung, C. G. (1960). The structure and dynamics of the psyche. London: Routledge and Keegan Paul.

* Kanthamani, H., & Broughton, R. S. (1992). An experiment in ganzfeld and dreams: A further confirmation. Proceedings of the 35th Annual Convention of the Parapsychological Association (pp. 59-73). Las Vegas, NV: Parapsychological Association.

* Kanthamani, H., & Khilji, A. (1990). An experiment in ganzfeld and dreams: A confirmatory study. Proceedings of the 33rd Annual Convention of the Parapsychological Association (pp. 126-137), Chevy Chase, MD: Parapsychological Association.

* Kanthamani, H., Khilji, A., & Rustomji-Kerns, R. (1988). An experiment in ganzfeld and dreams with a clairvoyance technique. In Proceedings of the 31st Annual Convention of the Parapsychological Association (pp. 412-423). Durham, NC: Parapsychological Association.

Kolata, G. B. (1981). Drug found to help heart attack survivors. Science, 214, 774-775.

Krippner, S. (1972). Concerning the scientific study of the human aura. Journal of the Society for Psychical Research, 46, 1-21.

Krippner, S. (1991). An experimental approach to the anomalous dream. In J. Gackenbach, & A. A. Sheikh (Eds.), Dream images: A call to mental arms (pp. 31-54). Amityville, NY: Baywood.

Krippner, S. (1993). The Maimonides ESP-dream studies. Journal of Parapsychology, 57, 39-54.

* Krippner, S., Honorton, C., & Ullman, M. (1972). A second precognitive dream study with Malcolm Bessent. Journal of the American Society for Psychical Research, 66, 269-279.

* Krippner, S., Honorton, C., & Ullman, M. (1973). An experiment in dream telepathy with "The Grateful Dead." Journal of the American Society of Psychosomatic Dentistry and Medicine, 20, 9-18.

* Krippner, S., Honorton, C., Ullman, M., Masters, R. E. L., & Houston, J. (1971). A long-distance 'sensory bombardment' study of ESP in dreams. Journal of the American Society for Psychical Research, 65, 468-475.

* Krippner, S., & Ullman, M. (1970). Telepathy and dreams: A controlled experiment with electroencephalogram-electro-oculogram monitoring. Journal of Nervous and Mental Disease, 151, 394-403.

* Krippner, S., Ullman, M., & Honorton, C. (1971). A precognitive dream study with a single subject. Journal of the American Society for Psychical Research, 65, 192-203.

Lowry, R. (2001-2014). VassarStats Binomial Probabilities. Retrieved May 20, 2014, from http://faculty.vassar.edu/lowry/binomialX.html.

Luke, D. (2002). A further exploratory investigation of dream precognition using consensus judging, dynamic targets and post judging target dramas. Paper presented at the 26th International Conference of the Society for Psychical Research, September 2002, Manchester Conference Centre.

* Luke, D., & Zychowicz, K. (2014). Working the graveyard shift at the witching hour: Further exploration of dreams, psi and circadian rhythms. International Journal of Dream Research, 7(2), 1-8.

Luke, D., Zychowicz, K., Richterova, O., Tjurina, I., & Polonnikova, J. (2010). An indirect approach to the neurobiology of psi: Precognition and circadian rhythms. Proceedings of presented papers to the 6th Psi Meeting: Neuroscience & Parapsychology, Curitiba, Brazil, 106-108.

* Luke, D., Zychowicz, K., Richterova, O., Tjurina, I., & Polonnikova, J. (2012). A sideways look at the neurobiology of psi: Precognition and circadian rhythms. NeuroQuantology, 10(3), 580-590.

* Markwick, B., & Beloff, J. (1983). Dream states and ESP: A distance experiment with a single subject. In W.G. Roll, J. Beloff, & R.A. White (Eds.), Research in parapsychology 1982 (pp. 228-230). New York: Scarecrow Press.

* Markwick, B., & Beloff, J. (1988). Dream states and ESP: A distance experiment featuring a pure clairvoyance, free-response design. In D. H. Weiner, & R. L. Morris (Eds.), Research in parapsychology 1987 (pp. 77-81). London: Scarecrow Press.

May, E. C., Utts, J. M., & Spottiswoode, J. P. (1995). Decision augmentation theory: Toward a model of anomalous mental phenomena. Journal of Parapsychology, 59, 195-220.

* McLaren, I., & Sargent, C. L. (1982). Awareness of success in free-response dream ESP testing with a single subject. In W.G. Roll, R.L. Morris, & R.A. White (Eds.), Research in parapsychology 1981 (pp. 195-196). Metuchen, NJ: Scarecrow Press.

Milton, J. (1997). Meta-analysis of free-response ESP studies without altered states of consciousness. Journal of Parapsychology, 61, 279-319.

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

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. Psychological Bulletin, 125(4), 387-391. doi:10.1037/0033-2909.125.4.387

Palmer, J. (2009). Decision augmentation in a computer guessing task. Journal of Parapsychology, 73, 119-136.

Parker, A. (1975). States of mind: Altered states of consciousness and ESP. New York: Taplinger.

Parker, A. (2005). Psi and altered states of consciousness. In M. A. Thalbourne & L. Storm (Eds.). (2005). Parapsychology in the twenty-first century: Essays on the future of psychical research (pp. 65-89). Jefferson, NC: McFarland.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. Psychological Bulletin, 112, 160-164.

Radin, D. (2006). Entangled minds: Extrasensory experiences in a quantum reality. New York: Paraview.

Rao, K. R. (2001). Basic research in parapsychology (2nd ed.). Jefferson, NC: McFarland.

Rechtschaffen, A. (1973). The psychophysiology of mental activity during sleep. In F. J. McGuigan, & R. A. Schoonover (Eds.), The psychophysiology of thinking (pp. 153-205). New York: Academic Press.

Rhine, L. E. (1981). The invisible picture: A study of psychic experiences. Jefferson, NC: McFarland.

Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., & Greenwood, J. A. (1940/1966). Extra-sensory perception after sixty years. Boston, MA: Bruce Humphries.

* Robinson, E. (2009). Student brief: Examining the case for dream precognition. European Journal of Parapsychology, 24, 93-110.

Rock, A. J., & Krippner, S. (2012). States of consciousness or states of phenomenology? In A. E. Cavanna & A. Nani (Eds.), Consciousness: States, mechanisms and disorders (pp. 55-65). Hauppauge, NY: Nova Science Publishers, Inc..

Roe, C. A., Davey, R., & Stevens, P. (2003). Are ESP and PK aspects of a unitary phenomenon? A preliminary test of the relationship between ESP and PK. Journal of Parapsychology, 67, 343-366.

* Roe, C. A., Jones, L., & Maddern, C. (2007). A preliminary test of the 'four factor model' using a dream ESP protocol. Journal of the Society for Psychical Research, 71, 35-42.

Roe, C. A., & Sherwood, S. J. (2009). Evidence for extrasensory perception in dream content: A review of experimental studies. In S. J. Krippner & D. Joffe-Ellis (eds.) Perchance to dream: The frontiers of dream psychology (pp. 211-240). New York: Nova Science Publishers.

Roe, C. A., Sherwood, S. J., & Farrell, L. (2007). Displacement in dream ESP research: Assessing the role of feedback during judging. Paper presented at the 31st International Conference of the Society for Psychical Research, Cardiff University, 31 Aug-2 Sept.

* Roe, C. A., Sherwood, S. J., Farrell, L., Savva, L., & Baker, I. (2007c). Assessing the roles of the sender and experimenter in dream ESP research. European Journal of Parapsychology, 22, 175-192.

* Roe C. A., Sherwood, S. J., Luke, D. P., & Farrell, L. M. (2002). An exploratory investigation of dream GESP using consensus judging and dynamic targets. Journal of the Society for Psychical Research, 66, 225-238.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.) (2005). Publication bias in meta-analysis: Prevention, assessment, and adjustments. Chichester, UK: Wiley.

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. Psychonomic Bulletin and Review, 18, 682-689. doi: 10.3758/s13423-011-0088-7

Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010). Psychological Bulletin, 139, 241–247. doi:10.1037/a0029008.

Sannwald, G. (1963). On the psychology of spontaneous paranormal phenomena. International Journal of Parapsychology, 5, 274-292.

* Sargent, C. L., & Harley, T. A. (1982). Precognition testing with free-response techniques in the ganzfeld and the dream state. European Journal of Parapsychology, 4, 243-256.

Schooler, J. (2011). Unpublished results hide the decline effect. Nature, 470(7335), 437. doi:10.1038/470437a.

* Sherwood, S. J., Dalton, K., Steinkamp, F., & Watt, C. (2000). Dream clairvoyance study II using dynamic video-clips: Investigation of consensus voting judging procedures and target emotionality. Dreaming, 10, 221-236.

Sherwood, S. J., & Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP studies. In J. Alcock, J. Burns, & A. Freeman (Eds.), Psi wars: Getting to grips with the paranormal (pp. 85-109). Thorverton, UK: Imprint Academic.

Sherwood, S. J., & Roe, C. A. (2013). An updated review of dream ESP studies conducted since the Maimonides dream ESP program. In S. Krippner, A. J. Rock, J. Beischel, & H. Friedman (Eds.), Advances in parapsychological research 9 (pp. 38-81). Jefferson, NC: McFarland.

* Sherwood, S. J., Roe, C. A., Simmonds, C. A., & Biles, C. (2002). An exploratory investigation of dream precognition using consensus judging and static targets. Journal of the Society for Psychical Research, 66, 22-28.

Spencer, B. (1995). Correlations, sample size, and practical significance: A comparison of selected psychological and medical investigations. Journal of Psychology, 129, 469-475.

Simmonds-Moore, C. A., & Moore, S. L. (2009). Exploring how gender role and boundary thinness relate to paranormal experiences, beliefs and performance on a forced-choice clairvoyance task. Journal of the Society for Psychical Research, 73, 129-143.

Stanford, R. G. (1977). Conceptual frameworks of contemporary psi research. In B. B. Wolman (Ed.), Handbook of parapsychology (pp. 823-858). New York: Van Nostrand Reinhold.

Stanford, R. G. (1978). Towards reinterpreting psi events. Journal of the American Society for Psychical Research, 72, 197-214.

Steering Committee of the Physicians Health Study Research Group (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. New England Journal of Medicine, 318, 261-264.

Steinkamp, F. (2000). Acting on the future: A survey of precognitive experiences. Journal of the American Society for Psychical Research, 94, 37-59.

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

Storm, L. (2009). Investigations of the I Ching: II. Reliability and validity studies. Australian Journal of Parapsychology, 9, 111-142.

Storm, L., & Ertel, S. (2001). Does psi exist? Comments on Milton and Wiseman's (1999) meta-analysis of ganzfeld research. Psychological Bulletin, 127, 424-433. doi:10.1037/0033-2909.127.3.424

Storm, L., Ertel, S., & Rock, A. J. (2013). The sheep-goat effect as a matter of compliance vs. noncompliance: The effect of reactance in a forced-choice ball selection test. Journal of Scientific Exploration, 27(3), 393-411.

Storm, L., & Rock, A. J. (2015). Dreaming of psi: A narrative review and meta-analysis of dream-ESP studies at the Maimonides Dream Laboratory and beyond. In J. Davies & D. Pitchford (Eds.), Stanley Krippner: A life of dreams, myths, and visions (pp. 117-137). San Francisco, CA: University Professors Press.

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

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analysis of free-response studies, 1992-2008: Assessing the noise reduction model in parapsychology. Psychological Bulletin, 136, 471- 485.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analyses of ESP studies 1987-2008: Assessing the success of the forced-choice design in parapsychology. Journal of Parapsychology, 76, 243-273.

Storm, L., Tressoldi, P. E., & Utts, J. (2013). Testing the Storm et al. (2010) meta-analysis using Bayesian and Frequentist approaches: Reply to Rouder et al. (2013). Psychological Bulletin, 139(1), 248-254. doi:10.1037/a0029506.

Strauch, I. (1970). Dreams and psi in the laboratory. In R. Cavanna (Ed.), Psi favorable states of consciousness (pp. 46-54). New York, NY: Parapsychology Foundation.

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

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

Thouless, R. H., & Wiesner, B. P. (1947). The psi process in normal and "paranormal" psychology. Proceedings of the Society for Psychical Research, 48, 177-196.

Timm, U. (2000). The problem of finding an optimal measure of effect size for psi experiments. Journal of Parapsychology, 64, 253–254.

Tressoldi, P. E. (2011). Extraordinary claims require extraordinary evidence: The case of non-local perception, a classical and Bayesian review of evidences. Frontiers in Psychology, 2, 117. doi:10.3389/fpsyg.2011.00117

* Ullman, M. (1969). Telepathy and dreams. Experimental Medicine & Surgery, 27, 19-38.

* Ullman, M., & Krippner, S. (1969). A laboratory approach to the nocturnal dimension of paranormal experience: Report of a confirmatory study using the REM monitoring technique. Biological Psychiatry, 1, 259-270.

* Ullman, M., Krippner, S., & Feldstein, S. (1966). Experimentally-induced telepathic dreams: Two studies using EEG-REM monitoring technique. International Journal of Parapsychology, 2, 420-437.

* Ullman, M., & Krippner, S. with Vaughan, A. (1973). Dream telepathy: Experiments in nocturnal ESP. Jefferson, NC: McFarland.

Ullman, M., & Krippner, S. with Vaughan, A. (1989). Dream telepathy: Experiments in nocturnal ESP (2nd ed.). Jefferson, NC: McFarland.

Ullman, M., & Krippner, S. with Vaughan, A. (2002). Dream telepathy: Experiments in nocturnal extrasensory perception (3rd ed.). Charlottesville, VA: Hampton Roads.

Utts, J., Norris, M., Suess, E., & Johnson, W. (2010). The strength of evidence versus the power of belief: Are we all Bayesians? In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society: Proceedings of the Eighth International Conference on Teaching Statistics. Voorburg, the Netherlands: International Statistical Institute.

* Van de Castle, R. L. (1971). The study of GESP in a group setting by means of dreams. Journal of Parapsychology, 35, 312.

Van de Castle, R. L. (1977). Sleep and dreams. In B.B. Wolman (Ed.) Handbook of parapsychology (pp. 473-499). New York: Van Nostrand Reinhold.

Van de Castle, R. L. (2009). Dreams and ESP. In C. Roe, W. Kramer, & L. Coly (Ed.), Proceedings of an international confer¬ence: Utrecht II: Charting the future of parapsychology (2008) (pp. 72-114). Parapsychology Foundation, Inc.

Vaughan, A., & Houck, J. (1993). A "success" test of precognition and attitude toward the future. Journal of the Society for Psychical Research, 59, 259-267.

Vaughan, A., & Houck, J. (2000). Intuition-training software: A second pilot study. Journal of the Society for Psychical Research, 63, 177-184.

Von Lucadou, W. (2001). Hans in luck: The currency of evidence in parapsychology. Journal of Parapsychology, 65, 3-16.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). Journal of Personality and Social Psychology, 100, 426-432. doi:10.1037/a0022790.

Watt, C. (1996). What makes a good psi target? Three studies of forced-choice ESP varying target emotionality and complexity. Journal of Parapsychology, 60, 25-41.

* Watt, C. (2014). Precognitive dreaming: Investigating anomalous cognition and psychological factors. Journal of Parapsychology, 78, 115-125.

Watt, C., & Valášek, M. (2015). Postscript to Watt (2014) on precognitive dreaming: Investigating anomalous cognition and psychological factors. Journal of Parapsychology, 79, 105-107.

Watt, C., Wiseman, R. & Vuillaume, L. (2015). Dream precognition and sensory incorporation: A controlled sleep laboratory study. Journal of Consciousness Studies, 22, 172-190.

* Weiner, D. H., & McCain, D. L. (1981). The use of the majority-vote technique in dream ESP: A pilot study. Journal of Parapsychology, 45, 156-157.

## Appendix A

*Table A1*. Dream-ESP Studies by Author, Lab (MDL = Maimonides), Study Type, Trials, Hits, *Z* Scores, and Effect Size Values.

| No. | Study | Author | MDL[a] | REM[b] | Type of ESP[c] | Trials | Hits | *z* | *ES* (z/√n) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Belvedere & Foulkes (1971) | 0 | 1 | TE | 8 | 3 | -0.35 | -0.122 |
| 2 | 2a | Braud (1977)—Pilot | 0 | 0 | TE | 50 | 3 | -6.08 | -0.860 |
| 3 | 2b | Braud (1977)—Experiment 1 | 0 | 0 | TE | 30 | 19 | 1.28 | 0.234 |
| 4 | 2c | Braud (1977)—Experiment 2 | 0 | 0 | TE | 36 | 19 | 0.17 | 0.028 |
| 5 | 3a | Child et al. (1977)—Experiment 1[d] | 0 | 0 | TE | 8 | n/a | 1.54 | 0.543 |
| 6 | 3b | Child et al. (1977)—Experiment 2[d] | 0 | 0 | TE | 7 | n/a | 1.90 | 0.718 |
| 7 | 4 | Dalton et al. (1999) | 0 | 0 | CL | 32 | 15 | 2.65 | 0.468 |
| 8 | 5 | Dalton et al. (2000) | 0 | 0 | CL | 16 | 7 | 1.41 | 0.353 |
| 9 | 6 | Foulkes et al. (1972) | 0 | 1 | TE | 8 | 5 | 0.35 | 0.124 |
| 10 | 7 | Harley (1989) | 0 | 0 | CL | 20 | 2 | -1.29 | -0.288 |
| 11 | 8 | Hearne (1981a) | 0 | 1 | TE | 2 | 0 | 0.00 | 0.000 |
| 12 | 9a | Hearne (1981b)—slow wave sleep[e] | 0 | 0 | TE | 8 | n/a | 0.58 | 0.205 |
| 13 | 9b | Hearne (1981b)—REM[e] | 0 | 1 | TE | 8 | n/a | 0.04 | 0.014 |
| 14 | 10 | Hearne (1987) | 0 | 1 | TE | 8 | 3 | -0.35 | -0.122 |
| 15 | 11 | Hearne (1989) | 0 | 0 | TE | 10 | 6 | 0.32 | 0.101 |
| 16 | 12a | Hearne & Worsley (1977)—close[f] | 0 | 1 | TE | 4 | n/a | 0.82 | 0.409 |
| 17 | 12b | Hearne & Worsley (1977)—not close[f] | 0 | 1 | TE | 4 | n/a | 0.67 | 0.336 |
| 18 | 13 | Honorton et al. (1972)—Vaughan, Harris, Parise | 1 | 1 | TE | 203 | 105 | 0.42 | 0.029 |
| 19 | 14 | Kanthamani & Broughton (1992)[d] | 0 | 0 | CL | 20 | n/a | 3.05 | 0.682 |
| 20 | 15 | Kanthamani & Khilji (1990)[d] | 0 | 0 | CL | 20 | n/a | 1.70 | 0.380 |
| 21 | 16a | Kanthamani et al. (1988)—Preliminary | 0 | 0 | CL | 4 | n/a | -0.45 | -0.225 |
| 22 | 16b | Kanthamani et al. (1988)—Pilot | 0 | 0 | CL | 10 | n/a | 0.26 | 0.082 |
| 23 | 17 | Krippner et al. (1972)—2nd Bessent | 1 | 1 | PR | 8 | 7 | 1.81 | 0.640 |
| 24 | 18 | Krippner et al. (1973)—Grateful Dead | 1 | 1 | TE | 12 | 7 | 0.29 | 0.084 |
| 25 | 19 | Krippner et al. (1971)—Sensory Bombardment | 1 | 1 | TE | 8 | 8 | 2.66 | 0.940 |
| 26 | 20 | Krippner & Ullman (1970)—Van de Castle | 1 | 1 | TE | 8 | 6 | 1.06 | 0.375 |
| 27 | 21 | Krippner et al. (1971)—1st Bessent | 1 | 1 | PR | 8 | 7 | 1.81 | 0.640 |
| 28 | 22 | Luke & Zychowicz (2014)—3am/8am[g] | 0 | 0 | PR | 268 | 69 | 0.21 | 0.013 |
| 29 | 23 | Luke et al. (2012)—3am/8am[g] | 0 | 0 | PR | 143 | 33 | -0.43 | -0.036 |
| 30 | 24 | Markwick & Beloff (1983) | 0 | 0 | CL/TE | 100 | 25 | 1.13 | 0.113 |
| 31 | 25 | Markwick & Beloff (1988) | 0 | 0 | CL/PR | 100 | 22 | 0.38 | 0.038 |
| 32 | 26 | McLaren & Sargent (1982) | 0 | 0 | PR | 17 | 1 | -1.64 | -0.398 |
| 33 | 27 | Robinson (2009) | 0 | 0 | PR | 100 | 52 | 0.30 | 0.030 |
| 34 | 28 | Roe, Jones, & Maddern (2007) | 0 | 0 | CL | 15 | 2 | -1.37 | -0.354 |
| 35 | 29a | Roe et al. (2007)—Sender | 0 | 0 | TE | 40 | 12 | 0.55 | 0.087 |
| 36 | 29b | Roe et al. (2007)—No Sender | 0 | 0 | CL | 40 | 14 | 1.28 | 0.202 |
| 37 | 30 | Roe, Sherwood, Luke, & Farrell (2002) | 0 | 1 | CL | 31 | 9 | 0.31 | 0.056 |
| 38 | 31 | Sargent & Harley (1982) | 0 | 0 | PR | 20 | 8 | 1.29 | 0.288 |
| 39 | 32 | Sherwood et al. (2000) | 0 | 0 | CL | 28 | 11 | 1.53 | 0.289 |
| 40 | 33 | Sherwood et al. (2002) | 0 | 0 | PR | 12 | 2 | -0.28 | -0.081 |

(continued)

*Table A1*. Dream-ESP Studies by ... (continued)

| No. | Study | Author | MDL[a] | REM[b] | Type of ESP[c] | Trials | Hits | *z* | *ES* (z/√n) |
|---|---|---|---|---|---|---|---|---|---|
| 41 | 34a | Ullman (1969)—2nd screening | 1 | 1 | TE | 12 | 4 | -0.87 | -0.251 |
| 42 | 34b | Ullman (1969)—Posin | 1 | 1 | TE | 8 | 6 | 1.06 | 0.375 |
| 43 | 35 | Ullman & Krippner (1969)—2nd Erwin | 1 | 1 | TE | 8 | 8 | 2.66 | 0.940 |
| 44 | 36a | Ullman et al. (1966)—1st screening | 1 | 1 | TE | 12 | 7 | 0.29 | 0.084 |
| 45 | 36b | Ullman et al. (1966)—1st Erwin | 1 | 1 | TE | 7 | 5 | 0.75 | 0.283 |
| 46 | 37a | Ullman et al. (1973)—Grayeb | 1 | 1 | TE | 8 | 3 | -0.35 | -0.124 |
| 47 | 37b | Ullman et al. (1973)—Pilot Sessions (H) | 1 | 1 | TE | 67 | 53 | 4.64 | 0.567 |
| 48 | 37c | Ullman et al. (1973)—Pilot Sessions (K) | 1 | 1 | PR | 2 | 2 | 0.67 | 0.474 |
| 49 | 37d | Ullman et al. (1973)—Pilot Sessions (N) | 1 | 1 | CL | 8 | 5 | 0.35 | 0.124 |
| 50 | 38 | Van de Castle (1971) | 0 | 0 | TE | 150 | 95 | 3.18 | 0.260 |
| 51 | 39 | Watt (2014) | 0 | 0 | PR | 200 | 64 | 2.20 | 0.156 |
| 52 | 40 | Weiner & McCain (1981)[d] | 0 | 0 | CL | 12 | n/a | 2.03 | 0.586 |

Note. [a] 1 = MDL, 0 = non-MDL; [b] 1 = REM, 0 = Non-REM; [c] TE = telepathy, CL = clairvoyance, PR = precognition; [d] original t score converted to *z* score (see Appendix B); [e] *z* score derived from *F* values (see Appendix B); [f] *z* score derived from Sandler's A values (see Appendix B); [g] recalled-dream data only (the Proceedings paper by Luke et al., 2010, is listed in Sherwood & Roe's, 2013, pp. 50-51, table, but it is the "same in content [as Luke et al., 2012] but different in title only"—D. Luke, personal communication, November 26, 2013).

## Appendix B

**Calculations of Statistics Used in the Analyses**

*Calculation of Z Scores and Effect Sizes*
All *z* scores for studies where $n \leq 1000$ (where $n$ = number of trials) were calculated from Exact Binomial *P* values (Source: http://faculty.vassar.edu/lowry/binomialX.html):

$$P(k \text{ out of } n) = \frac{n!}{k!(n-k)!} * (p^k)(q^n - k)$$

Formula 1

where $n$ = number of trials, $k$ = number of hits, $p$ = probability of a hit, and $q$ = probability of no hit (i.e., $1 - p$).

If $np \geq 5$ and $nq \geq 5$, binomial probabilities were estimated by way of the binomial approximation of the normal distribution, according to the formula:

$$z = \frac{(k - M) \pm .5}{\sigma}$$

Formula 2

where $M = np$ (the mean of the binomial sampling distribution), and $\sigma = \sqrt{[npq]}$ (the *SD* of the binomial sampling distribution). NB: This formula includes a continuity correction (± .5) that yields negative *z* scores for chance scoring. One can see that repeated use of Formula 2, given the appropriate data, will yield conservative outcomes in meta-analyses because the mean *z* and *ES* values are consistently nudged in a negative direction. In some cases (studies 1, 5, 6, 10, 17, 33, 34b, 35, 36b, 37a, and 37c), *z* scores were determined from calculated *p* values. For studies 3a, 3b, 15, and 40, in Table A1, *t* values were converted to *z* scores using the formula:

$$z = \sqrt{[df * \log_e(1 + t^2/df)]} \times \sqrt{[1 - 1/(2*df)]}$$

Formula 3

Thus,

- Child, Kanthamani & Sweeney (1977)—Expt 1, $t(7) = 1.87$, $z = 1.54$
- Child, Kanthamani & Sweeney (1977)—Expt 2, $t(4) = 2.69$, $z = 1.90$
- Kanthamani & Broughton (1992), $t(19) = 3.52$, $z = 3.05$
- Kanthamani & Khilji (1990), $t(19) = 1.79$, $z = 1.70$
- Weiner & McCain (1981) , $t(11) = 2.30$, $z = 2.03$

For the slow-wave sleep (SWS) group and the REM group in the study by Hearne (1981b), we converted *F* values, SWS: $F(1, 7) = .372$ and REM: $F(1, 7) = .002$, to *t* values where $t = \sqrt{F}$, so that $t = 0.61$, and $t = 0.04$, respectively. Using Formula 3 above, we then converted the *t* values to *z* scores ($z = 0.58$, and $z = 0.04$, respectively).

Effect size (*ES* or *r*) calculations were made using the formula:

$$r = z/\sqrt{n}$$

Formula 4

For the two "groups" in the study by Hearne and Worsley (1977), we converted the Sandler's A values ("Close: A = 1.09, not close: A = 1.67, with 3 dfs", p. 437) to *t* values, where:

$$t = \sqrt{df}/\sqrt{(A*df)}$$

Formula 5

so that $t = 0.96$, and $t = 0.77$, respectively. Using Formula 3 above, we then converted the *t* values to *z* scores ($z = 0.82$, and $z = 0.67$, respectively).

We use $z/\sqrt{n}$ mainly due to the 'simplicity' precedent set by Honorton and Ferrari (1989, p. 283).