



From Séances to Science: Investigating Psychic Phenomena with Statistics

Jessica Utts

Department of Statistics

University of California, Irvine

<http://www.ics.uci.edu/~jutts>

jutts@uci.edu



A Question for You

What do you really know to be true, and how do you know it?



How Do We Know What We Know?

- Private knowing (individual):
 - Experience
 - Gut feeling/intuition/faith
 - Belief in experts
- Public knowledge (shared):
 - Shared experience and belief
 - Physical and biological “laws”
 - Studies based on statistical analysis



Psi/Psychic/ESP/Anomalous Cognition

Having information that could not have been gained through the known senses.

- **Telepathy**: Info from another **person**
- **Clairvoyance**: Info from another **place**
- **Precognition**: Info from the **future**
- **Correlation**: Simultaneous access to info

For **proof** -> Source **isn't** important.

For **explanation** -> Source **is** important.



COINCIDENCE?

- I was visiting a friend when she received a call announcing a baby boy had been born to her brother and sister-in-law in Pennsylvania. I asked what they had named him, no name yet.
- Names "Timothy" and "Michael" popped into my head so strongly, that I said "If they name him Timothy or Michael, let me know."
- A few days later, my friend told me baby had been named "Timothy James." I reminded her of my prediction, which she reluctantly acknowledged, but she said "You were only half right." A few days later, in talking with her brother, she discovered otherwise. Timothy and Michael were the two names they had been trying to choose between.



THE STORY OF TIMOTHY AND MICHAEL

By chance alone what is the probability of this event?

- *Suppose there were about 100 reasonable names and I guessed two of them at random.*

The probability that I would guess:

Timothy and Michael is $(1/100) \times (1/99) = 1/9900$

Michael and Timothy is $(1/100) \times (1/99) = 1/9900$

Total = $2/9900$

- But what about equally impressive guesses:

Timothy and James $1/9900$

James and Timothy $1/9900$

Overall Total = $4/9900 = 1/2475$



Is That Amazing or Not?

- 2,475 days = about 6 years and 9 months...
- If I did something like this every day, I would be right about once every 7 years, by chance. So was I right by chance?? That seems unlikely, but:
- The proper question to ask is *not*:
By chance alone what is the probability of this unusual event?
- But rather:
By chance alone, what is the probability of something unusual happening to each of us occasionally?
Answer: Very high. So, this is *not* good evidence of psi.



HOW SCIENCE WORKS

Moving from anecdotes to knowledge:

- Create testable hypotheses
- Design suitable experiments to test them
- Analyze results, see if they support hypotheses
- Possibly create new testable hypotheses

And eventually...

- Accumulate sufficient [statistical] evidence, and hopefully, an explanation of how and why something works.



Controlled Experiments to Test Psychic Abilities

Crucial elements:

1. Safeguards to rule out cheating or ordinary means of communication
2. Knowledge of probabilities of various outcomes by chance alone

Examples... are these okay?

1. I am thinking of a number from 1 to 5. Guess it.
2. My assistant in California has shuffled a deck of cards (well!) and picked one. What suit is it?

Note: These are examples of *forced choice* experiments



Types of Experiments to Test Psychic Abilities

- **Forced Choice** experiments are like multiple choice tests.
 - Easy to analyze statistically
 - Poor way to generate good data because there is a “signal to noise” problem.
- **Free response** experiments are more like essay exams. Two types of experiments:
 - *Remote Viewing*, originally done by US Government
 - *Ganzfeld*, developed for other experiments in psychology and adapted for testing psychic abilities



US Government Remote Viewing Program

- 1972: Hal Puthoff at Stanford Research Institute (now SRI) conducted \$50,000 8-month project at the request of the CIA; invented “remote viewing”; results convinced CIA to continue.
- 1973: Project “Scanate” (Scanning by coordinate) test of a classified site in West Virginia.
 - *“One subject drew a detailed map of the building and grounds layout, the other provided information about the interior, including codewords, data subsequently verified by sponsor sources.” [Puthoff and Targ, 1975, recently declassified report.]*
 - Site in USSR equally good. Results seemed too good to be chance - CIA got very interested (and worried).



History of the US Government Remote Viewing Program, continued...

- 1973-75: More tests by CIA and others; publication of a book - *Mind Reach* (1977) by Targ and Puthoff. Severe and sometimes erroneous criticisms from extreme skeptics, led to “no publicity” policy at SRI.
- 1975-95: DIA and others funded research at SRI/SAIC.
- Later research focused on *process* questions, like if sender is needed (no), if hypnosis helps (no), if feedback is needed (probably).
- Classified “**operational**” work began at Fort Meade (“psychic spies”); mostly still classified.
- 1987-89: I was a Visiting Scientist in the program at SRI; I continued as a consultant until program ended



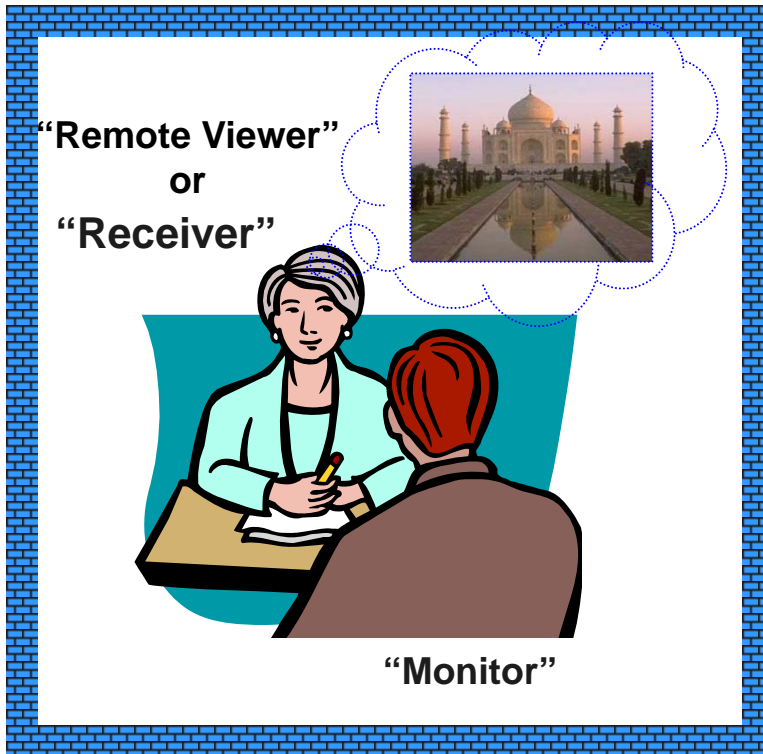
History of Remote Viewing, continued

- 1995: By a “Congressionally Directed Action” Congress asked CIA to evaluate and possibly take over program.
- Led to “Stargate” review by American Institutes for Research (AIR) – Utts, Hyman and AIR staff. (More later!)
- Nov 28, 1995: ABC News Nightline ran story, followed by Larry King and others. CIA released the AIR report. Government remote viewing program terminated.
- Today: Many of the former government remote viewers now teach training classes, some have written books. But some of the past work remains classified.
- Google search on “remote viewing” produces more than 1,700,000 hits

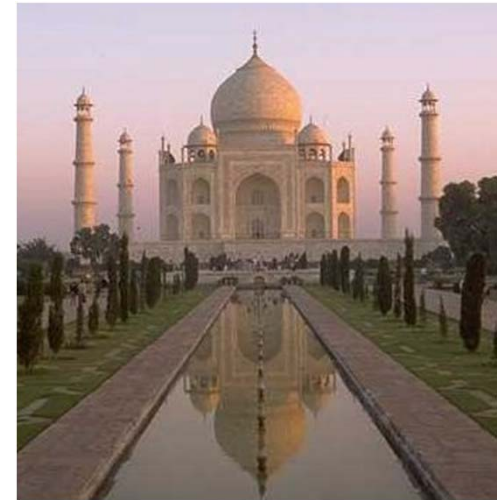
Remote Viewing Protocol

Special thanks to Dr. Edwin May for this and other SRI slides

10:00



15 Minutes



10:05





Target Material

- In the early experiments, physical locations were used and an “outbound experimenter” drove to a location, randomly selected from a set of possibilities.
- Later experiments used photographs, for example, randomly chosen from a set of 200 from *National Geographic* magazine.
- Short segments from movies sometimes used.
- Numbers, words, etc. are *not* used. (Signal to noise problem.)



Some Additional Details

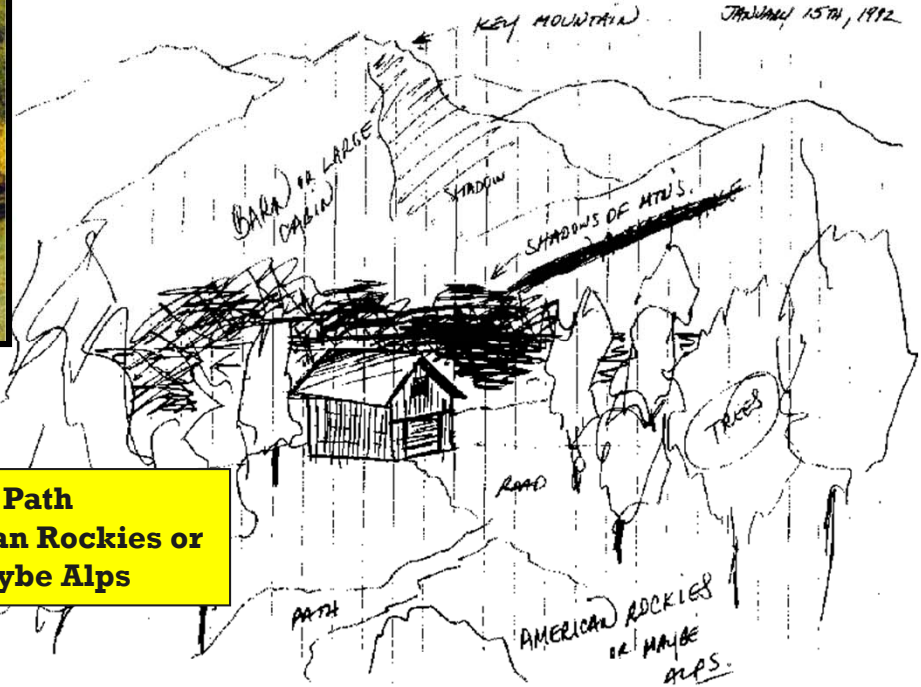
- After the session, the drawings and descriptions are copied and the original locked away.
- Feedback to the remote viewer is given by showing him/her the copy of what (s)he drew, along with the actual target photo or video.
- In some labs, the viewer is the judge and feedback isn't given until after judging. In others there is an independent judge.
- Meets **condition #1**: Safeguards to rule out cheating or ordinary means of communication

Example of an Excellent Match

(Experiment at SAIC/Stanford)



**Key Mountain
Barn or Large Cabin
Shadow
Shadows of Mtns.
Trees
Road**



**Path
American Rockies or
Maybe Alps**

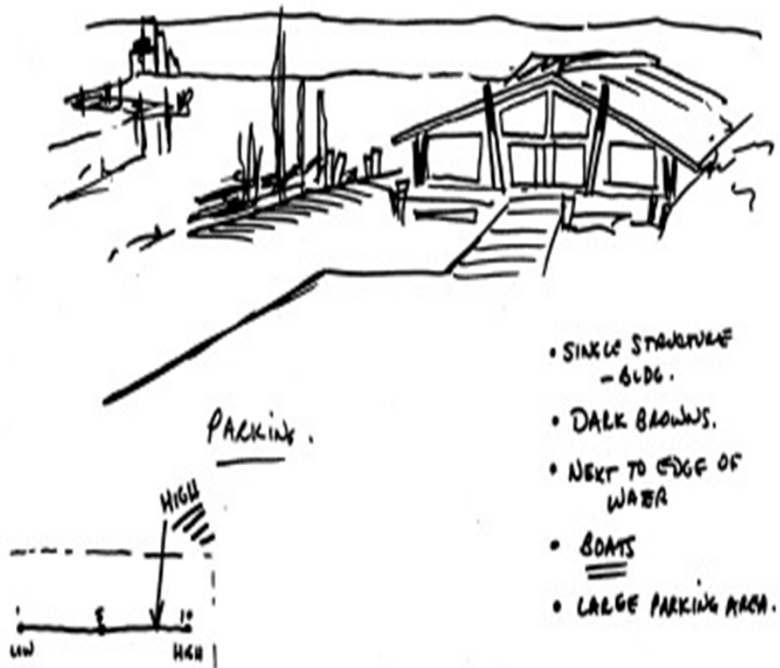
Early Remote Viewing Example (SRI)



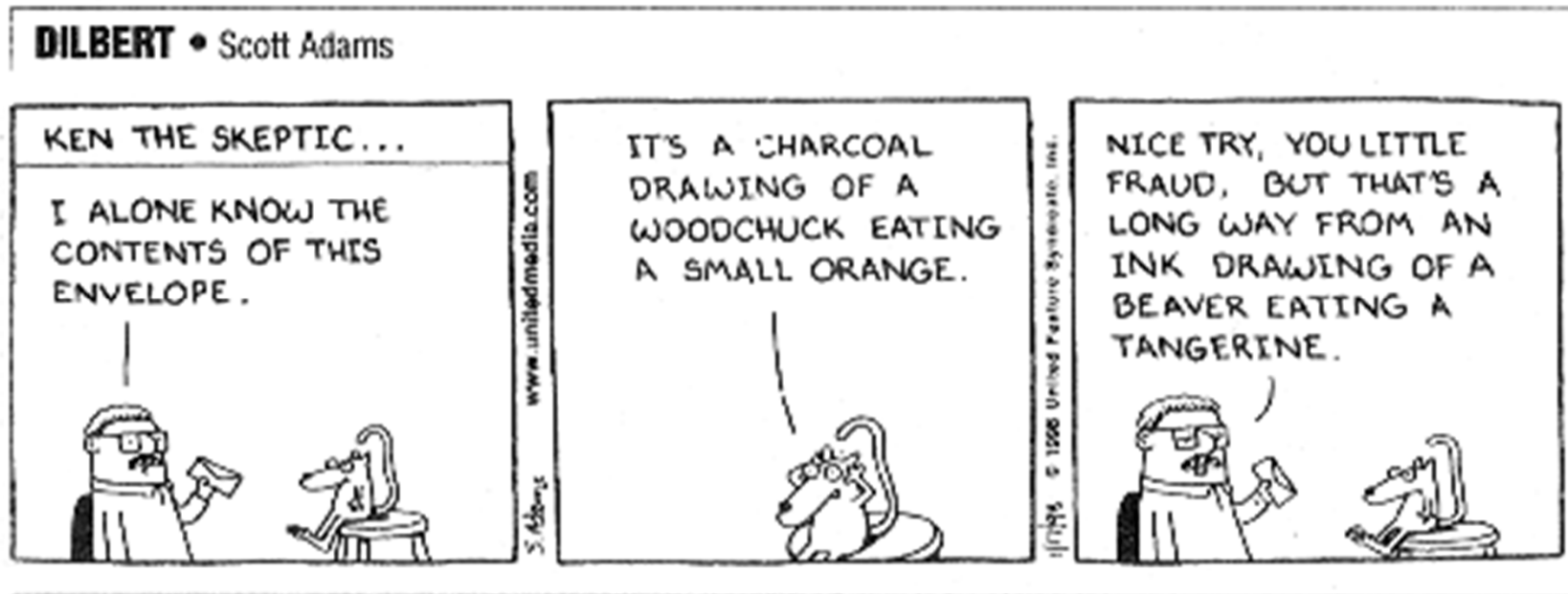
Target: Pete's Harbor Restaurant



How to Judge?



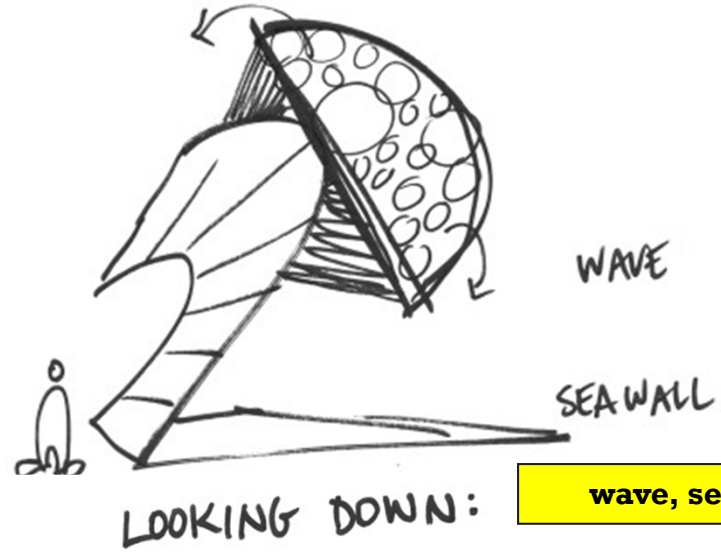
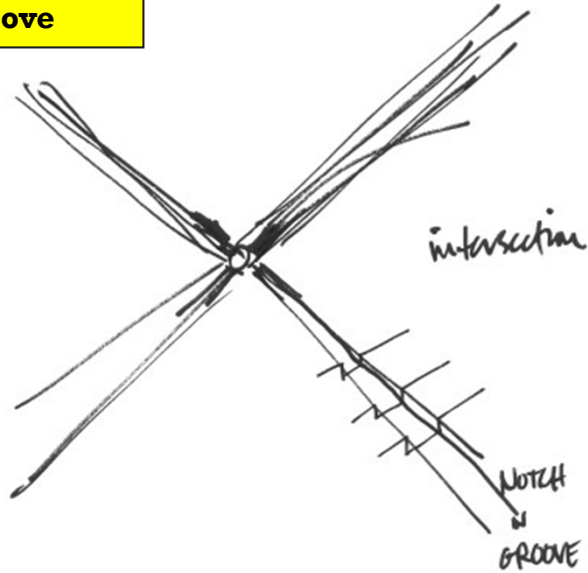
How NOT to Judge the Response



Can't use subjective probability of match
– too much room for personal bias.

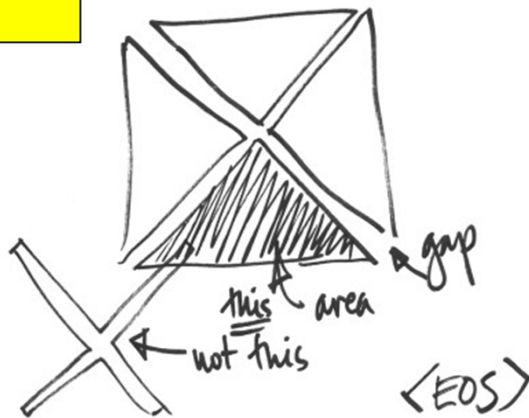
Typical Response – Novice

intersection,
notch, groove

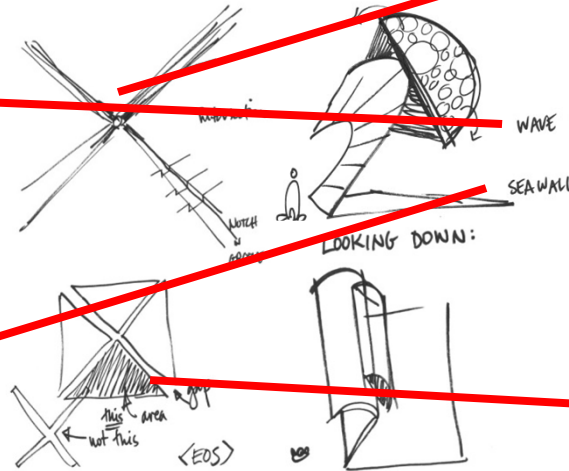
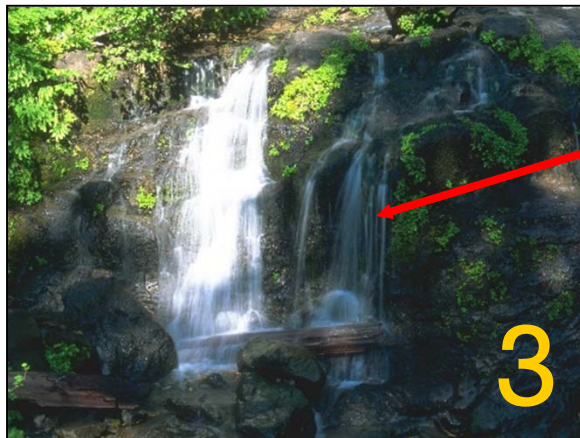
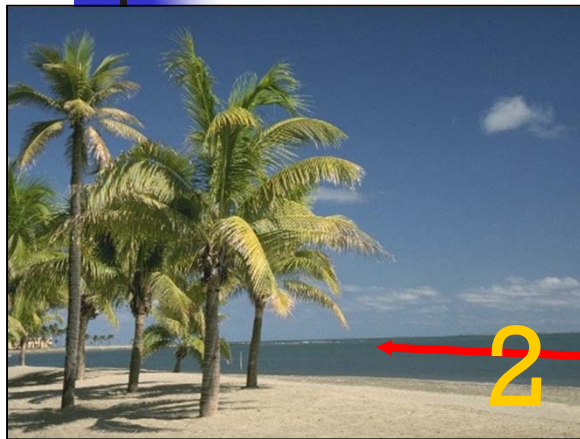


wave, sea wall

gap



Rank-Order or Direct Hit Judging



Simplest analysis just counts a “direct hit” if actual target is ranked #1. This example was scored as a direct hit.



An **Experiment** has many **Sessions**

- Before the *experiment*, a “target pool” is created - many packs of 4 dissimilar sets of photos (or short videos).
- Before each *session* begins a pack of 4 is randomly selected, then target within it (e.g. windmills). The *session* takes place, producing a response.
- After the *session*, a judge is given the response and the 4 choices from that target pack. Judge must assign the 4 ranks (and is of course blind to correct answer).
- For each *session*, result = the *rank* assigned to correct target, *or* “direct hit” if it gets 1st place rank. In some labs judge picks best match only (no ranks).



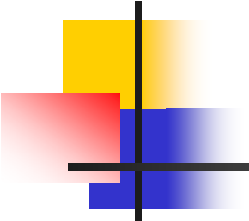
Experiments, Sessions, Probability

- Summary statistic for entire *experiment* (many *sessions*):

- Sum or ranks, or
- Number of direct hits

- Meets **Condition #2**:

Knowledge of **probabilities** of various outcomes by chance alone. For example, **probability of direct hit** = $1/4$.



Automated Ganzfeld Experiments Similar to Remote Viewing

- **Sender, receiver, experimenter.** Target selection same as remote viewing (random, packs of 4 photos or videos)
- **Sender** in sound-shielded room, looking at target on a TV screen.
- **Receiver** in sound isolation room with red light in eyes, white noise in ears, comfy chair.
- **Receiver** listens to relaxation tape. Then talks into microphone, attempting to describe the unknown target.

Ganzfeld Experiments, continued

- Experimenter and sender listen as receiver talks.
- Then *receiver* judges response. Shown 4 choices: the actual target and 3 decoys.
- Direct hit analysis usually used.
- Probability of a match by chance alone is $\frac{1}{4}$ or 25%.



Source:
<http://hopelive.hope.ac.uk/psychology/para/research.html>

What Constitutes Evidence from Statistical Studies?



- Small *p-values* (less than .05 is standard for concluding there is really something going on)
 - *p-value* = probability of observing results as extreme as those observed, **if** chance alone is the explanation. Small *p-value* = strong evidence against chance. (Similar to “odds against chance.”)
- *Confidence intervals* showing similar effects in a variety of similar situations, labs, etc.
 - A 95% confidence interval is an interval of values we are 95% confident covers the truth.
- Independent replication and meta-analysis
 - Remote viewing and ganzfeld, for instance



Simplest Model for RV and Ganzfeld

- Use direct hits only, p = probability of a correct match.
- Note that randomness is in selection of *target*, not in the response. People do not draw “randomly.”
- p = the probability that judge is able to pick the same *target* as the randomization does, *given* the response.
- By chance alone, $p = 1/4$.
- If we can verify that $p > 1/4$ over many experiments, it may indicate information was “received” from target, or some form of anomalous cognition occurred.



Summer 1995 Review of US Government Research Program

- 2 person team, Hyman and Utts, asked:
 - Does psychic functioning work?
 - Is it useful for intelligence work?
- 3 boxes of reports from government work at SRI and SAIC
- Told to focus on government program only, but I expanded the review to include other labs – replication is the hallmark of science!



P-value and C.I. Results of Early Free Response Experiments (for 1995 report for Congress)

Hit rates; remember with *four* choices **chance = 25%**

Government Studies in Remote Viewing:

- SRI International (1970's and 1980's)
966 trials, p-value = 4.3×10^{-11}
hit rate = 34%, 95% C.I. 31% to 37%
- SAIC (1990's)
455 trials, p-value = 5.7×10^{-7}
hit rate = 35%, C.I. 30% to 40%



Ganzfeld Results

- Early 80s Meta-analysis (some flaws identified, over-estimates truth)
492 trials, p-value = 6.5×10^{-12} , **hit rate = 38.1%**, C.I. 33.9% to 42.5%
- Psychophysical Research Laboratories, Princeton (1980's)
355 trials, p-value = .00005, **hit rate = 34.4%**, C.I. 29.4% to 39.6%
- University of Amsterdam, Netherlands (1990's)
124 trials, p-value = .0019, **hit rate = 37%**, C.I. 29% to 46%
- University of Edinburgh, Scotland (1990's)
97 trials, p-value = .0476, **hit rate = 33%**, C.I. 25% TO 44%
- Rhine Research Institute, North Carolina (1990's)
100 trials, p-value = .0446, **hit rate = 33%**, C.I. 24% to 42%



My Conclusion in AIR Report

- “Using the standards applied to any other area of science [that uses statistics], it is concluded that psychic functioning has been well established.”
(Reports available at <http://www.ics.uci.edu/~jutts>)
- Conclusion based on the above results (p-values and confidence intervals), and having visited several of the laboratories and seen their controls to eliminate cheating, etc.
- How do statisticians integrate research from many studies? Use “meta-analysis” to combine results.

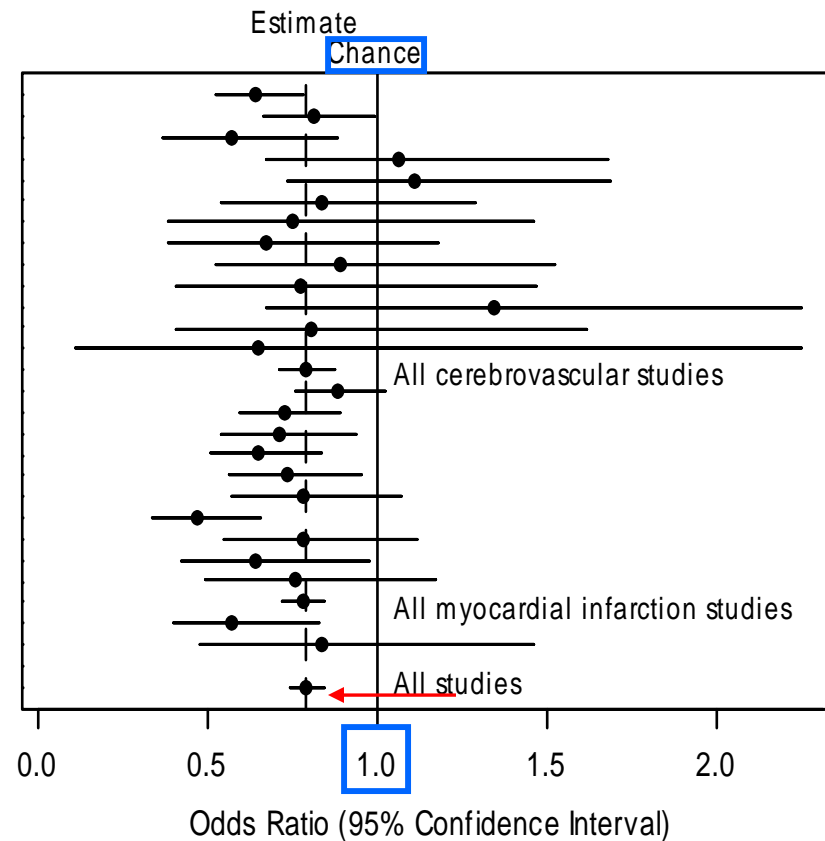


Meta-analysis: Non-controversial example
Aspirin and Recurring Vascular Disease
(*British Medical Journal*, summarized in *Science*)

- Meta-analysis of 25 clinical trials on recurrence of heart attack or stroke when taking aspirin versus placebo.
- Outcome of interest: Odds ratio
- Odds of recurrence aspirin/placebo
- Chance -> Odds ratio = 1
 - 25 Studies, 5 with *p-value* < .01
 - Combined odds ratio of 0.75, represents 25% drop in recurrence rate of heart attacks if taking aspirin

Confidence Intervals for Odds Ratio

Each line represents one study. Vertical lines at .75 (average for all studies) and 1 (value indicating no effect, just chance)



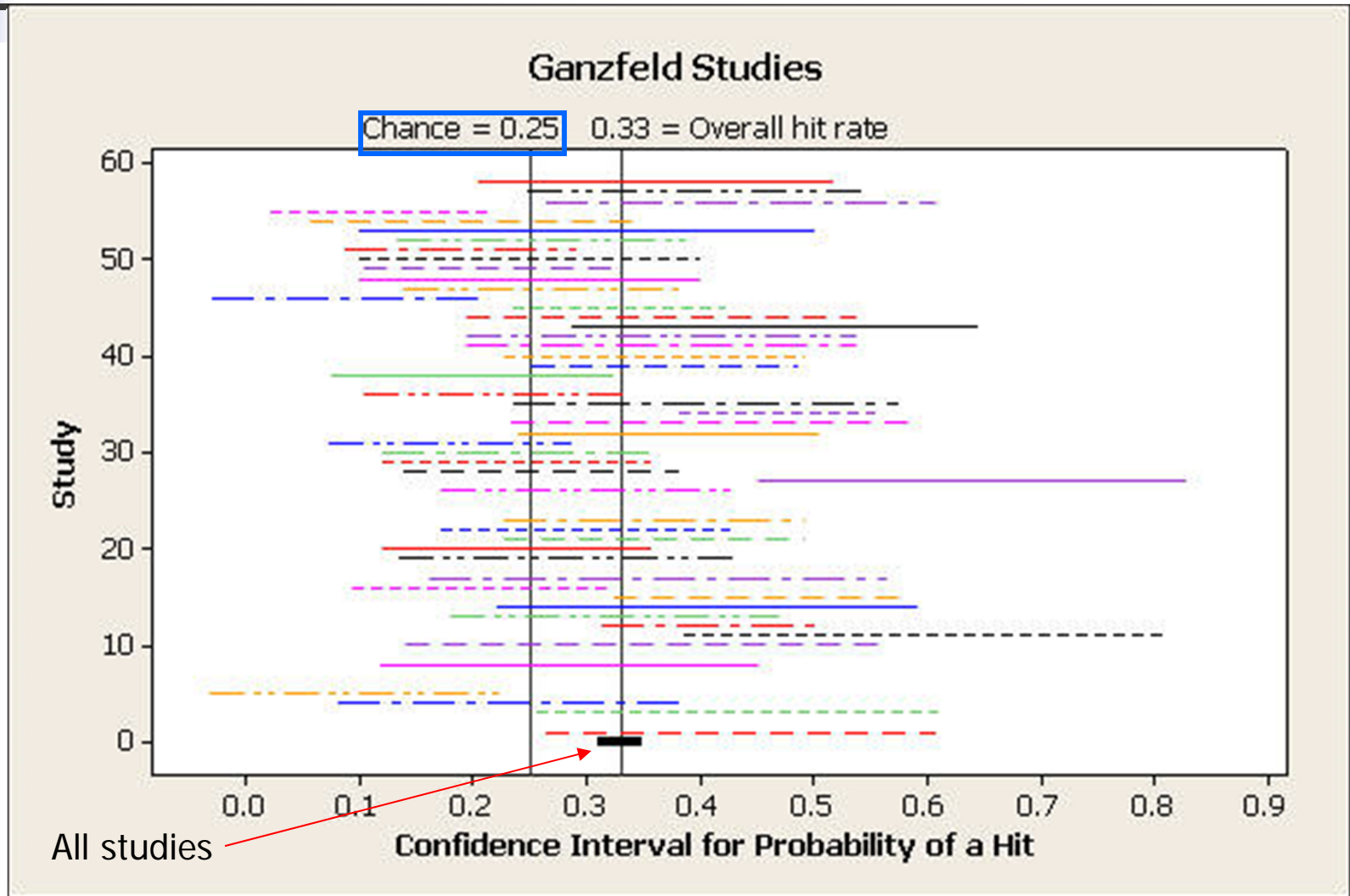


Simple meta-analysis for Ganzfeld

- 58 Studies from a variety of labs
- Results: 728 hits, $n=2206$, 33% hits
- 95% C.I. .31 to .35 (31% to 35%)
- $z=8.28$, p -value = 6.2×10^{-17}

Ganzfeld Studies

58 Studies, overall hit rate = 33%





Recent Update: Storm et al, Psych Bulletin

Meta-analysis of all new ganzfeld studies from 1997 to 2008

- 29 studies
- Total $n = 1498$ with 483 hits, 32.2% hit rate (25% expected by chance)
- 95% confidence interval is 29.9% to 34.6%
- Overall (exact binomial) p -value = 1.8×10^{-10}



Quotes about aspirin studies

- The trials were very **heterogeneous**, including a range of ages, a range of different diseases, a range of treatments, and so on.
- Though such risk reductions might be of some practical relevance, however, they are **surprisingly easy to miss**, even in some of the largest currently available clinical trials. If, for example, such an effect exists, then even if 2000 patients were randomized there would be an even chance of getting a false negative results...that is, of failing to achieve convincing levels of statistical significance ($p < .01$).
- The main results were obtained from the principal investigators in most cases. In some trials **the data obtained differed slightly from the data originally published**.
- The final meeting of collaborators was **supported** not only by the [UK] Medical Research Council and Imperial Cancer Research Fund but also by the **Aspirin Foundation**, Rhone-Poulenc Sante, Reckitt and Colman, **Bayer**, Eli Lilly, Beechams, and the United Kingdom Chest, Heart and Stroke Association.

And... what was to prevent having pill analyzed by local chemist?



Compare to Aspirin/Heart Attack Studies

- How are anomalous cognition (ac) - remote viewing and ganzfeld - results different from aspirin results?
 - If *same* standard applied, **ac results** are *stronger*. Cohen's *h* (measures effect) is .0875 (aspirin) vs .1767 (ganzfeld)
 - The **aspirin studies** had *more opportunity for fraud* and experimenter effects than did the ac studies.
 - The **aspirin studies** were *at least* as frequently **funded** and **conducted** by those with a *vested interest* in the outcome (aspirin companies).
 - Both used *heterogeneous* methods and participants.
- Why do many people believe the aspirin results, but either don't know about the anomalous cognition results or don't believe them?



Establishing Scientific Knowledge

- What roles do personal experience versus “objective” information play in what we think we know?
- Would *you* be more convinced by hundreds more statistical studies, or by one overwhelming personal experience?



Some Current Research

- Meta-analysis of another type of psi experiment, called “presentiment” in which physiology responds *before* an event happens
- Use “Bayesian statistics” to analyze results; combines prior beliefs with data to reach conclusions.
- Investigating possible models that would fit with what we know in physics