

PSYCHOLOGY MISCELLANY

No. 65 - November 2014

Kevin Brewer

ISSN: 1754-2200

Orsett Psychological Services
PO Box 179
Grays
Essex
RM16 3EW
UK

orsettpsychologicalservices@phonecoop.coop

This document is produced under two principles:

1. All work is sourced to the original authors. The images are all available in the public domain (most from http://commons.wikimedia.org/wiki/Main_Page). You are free to use this document, but, please, quote the source (Kevin Brewer 2014) and do not claim it as your own work.

This work is licensed under the Creative Commons Attribution (by) 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 171 2nd Street, Suite 300, San Francisco, California, 94105, USA.

2. Details of the author are included so that the level of expertise of the writer can be assessed. This compares to documents which are not named and it is not possible to tell if the writer has any knowledge about their subject.

Kevin Brewer BSocSc, MSc

An independent academic psychologist, based in England, who has written extensively on different areas of psychology with an emphasis on the critical stance towards traditional ideas.

A complete listing of his writings at <http://kmbpsychology.jottit.com>.

CONTENTS

	Page Number
1. KNOWING GOOD THERAPIES FROM BAD: AN EXAMPLE WITH NEURO-LINGUISTIC PROGRAMMING (NLP)	4
1.1. Distinguishing good therapies from bad	
1.1.1. Power therapies	
1.2. Neuro-Linguistic Programming	
1.3. Appendix 1A - Apparently getting better	
1.4. Appendix 1B - Thought Field Therapy (TFT)	
1.5. References	
2. REPLICATION AND THE "MANY LABS" PROJECT	16

1. KNOWING GOOD THERAPIES FROM BAD: AN EXAMPLE WITH NEURO-LINGUISTIC PROGRAMMING (NLP)

- 1.1. Distinguishing good therapies from bad
 - 1.1.1. Power therapies
- 1.2. Neuro-Linguistic Programming
- 1.3. Appendix 1A - Apparently getting better
- 1.4. Appendix 1B - Thought Field Therapy (TFT)
- 1.5. References

1.1. DISTINGUISHING GOOD THERAPIES FROM BAD

Norcross et al (2006) observed: "The burgeoning evidence-based practice (EBP) ¹ movement in mental health attempts to identify, implement, and disseminate treatments that have been proven demonstrably effective according to the empirical evidence ². This movement has provoked enormous controversy within organised psychology, and, with the exception of the general conviction that psychological practice should rely on empirical research, little consensus currently exists among the various stakeholders on either the decision rules to determine effectiveness or the treatments designated as 'evidence-based'" (p515).

This led Norcross et al (2006) to suggest that establishing what does not work (ie: discredited therapies and treatments ³) might be easier.

¹ Evidence is understood "as derived from systematic observation, sometimes over a long period of time. Often such evidence will be described as empirical or as evidence from observation; sometimes it will be described as evidence from experiment. Sometimes the observations have an underlying theoretical perspective, such as Darwin's observations leading to a theory of evolution..." (McQueen 2001 p263).

The Centers for Disease Control and Prevention in the USA in 2000 proposed a guide for establishing evidence in health care generally - "(i) suitability of evaluation design to attribute, with confidence, a change in an outcome caused by the given intervention; (ii) quality of study execution; (iii) numbers of studies; (iv) consistency of findings; (v) size of observed effects; and, in rare circumstances, (vi) expert opinion" (McQueen 2001 p264).

Writing in relation to healthcare, McQueen (2001) admitted to three problems when trying to establish universal evidence of effectiveness:

- i) The "rules of evidence" vary between disciplines;
- ii) No consensus over the "hierarchy of evidence" (ie: which type of evidence is "better" than another);
- iii) Complex treatments that make simple rules of evidence untenable.

Evidence-based practice is an idea with an inherent Western bias with the "reification of the positivist notion" (McQueen 2001).

² Individuals can spontaneously recover, and treatments involve the placebo effect, but "when psychotherapists base their interventions on the reliable research at their disposal, there is reason to believe that recipients will be helped - though not to the extent that many assume and all would wish" (Beyerstein 2001a p72).

³ Also called pseudoscientific, unvalidated, or "quack" (Norcross et al 2006).

Norcross et al (2006) collected 59 treatments and thirty assessment techniques used in mental health in the past one hundred years that were discredited - "unable to consistently generate treatment outcomes (treatment) or valid assessment data (tests) beyond that obtained by the passage of time alone, expectancy, base rates, or credible placebo" (p516). These were presented to 101 experts (randomly selected qualified mental health professionals in the USA)⁴, who reported their familiarity with the treatment or test, and rated each on a scale of (1) "not at all discredited" to (5) "certainly discredited".

The highest scoring treatments were "angel therapy for treatment of mental/behavioural disorders" and "use of pyramids for restoration of energy", which were both scored at 4.98. Fourteen treatments had scores over 4.5, which classed them as discredited by consensus, though familiarity with some was low (eg: colour therapy for treatment of mental/behavioural disorders - half of experts).

The assessment techniques received lower scores overall, with the highest scorer being "critical incident stress debriefing (CISD) for acute trauma" (3.25), and "psychosocial therapies for treatment of paedophilia" (3.23).

Significant differences in rating were found between experts based on their theoretical orientation (eg: cognitive-behavioural vs psychodynamic).

Some of the experts questioned the use of a single five-point scale for assessing the credibility of the treatment.

Norcross et al (2006) offered a note of caution: "Professional consensus does not equal an epistemic warrant; even experts can be and have been wrong"⁵. Expert opinions may be widely held because they are correct or because most experts simply share the same worldview. Test validity is conditional; clinical utility is purpose- and context-specific. A treatment or test considered discredited for one purpose might be credible for another. We should take care not to threaten innovation and creativity in psychological practice by branding all non-researched procedures as discredited. We must avoid hubris by remembering that contemporary treatments and tests may become discredited 30 years from now. We must avoid false pride, as science should be ever vigilant and self-correcting" (pp519-520).

Beyerstein (2001a) admitted that most therapies

⁴ Described as a Delphi poll or methodology. Individuals complete a questionnaire in the first round, and then complete it again (2nd round) after receiving feedback (eg: attitudes of others).

⁵ Dawes (1994) criticised "clinical judgment" or "clinical intuition" as being able to predict a client's behaviour (eg: recidivism or violence).

could be beneficial in one sense: "At the outset, it must be said that although fringe practices such as 'rebirthing' and Neuro-linguistic Programming are based on what Richard Rosen [1978] has aptly dubbed 'psychobabble', most of them probably do little damage in the long run – providing we overlook the costs of pandering to the narcissistic irrationalism of society's more affluent worriers. Despite their absurd premises, these therapeutic outliers at least provide clients of a certain metaphysical bent with comforting mythologies that explain why their lives are not as fulfilling as they had expected. Indirectly, these quaint rituals can supply existential support, emotional consolation, and even some useful spurs to change trouble-some habits. Thus, on balance, psychotherapies founded on ill-conceived assumptions may still prove beneficial if they furnish needed reassurance in an atmosphere where clients can mull over solutions to their dissatisfactions in life" (p70).

Beyerstein (2001a) highlighted three dangers of what he called "fringe therapies" - the risk of manipulation and fraud, untrained therapists who miss serious psychopathologies, and hardship (eg: encouraging clients' delusions of abduction by extraterrestrials).

Beyerstein (2001a) viewed psychoanalysis negatively because of problems in establishing its scientific basis and effectiveness. Furthermore, Beyerstein (2001a) blamed it for "perpetuating several popular misconceptions", namely:

i) Most psychological problems in adulthood stem from trauma in childhood - Not all traumatised children suffer psychologically as adults, and adults with non-trauma childhoods can have psychological problems.

ii) Individuals are inevitably damaged psychologically by negative events - Most individuals are not (eg: after a natural disaster many victims do not suffer from post-traumatic stress disorder).

iii) Unpleasant memories are repressed (ie: kept from the conscious mind) - Post-traumatic stress disorder suggests the opposite: that individuals cannot forget about unpleasant events even when they want to do so.

iv) The extremely traumatised mind splits into multiple personalities - The existence of multiple personality disorder (or dissociative identity disorder as now called) is disputed by many.

Beyerstein (2001a) listed the following as "questionable products in the therapeutic marketplace" -

aromatherapy, eye-movement desensitisation and reprocessing (EMDR) therapy, handwriting analysis (graphology), meditation as psychotherapy, and rebirthing-related techniques (ie: re-experiencing the womb and/or birth).

Beyerstein (2001a) was saddened: "As long as people refuse to think critically and to put psychotherapy methods to hard-nosed empirical tests, bogus treatments will continue to flood the market. It continues to amaze me that many people who demand extensive, impartial evaluations of automobiles or televisions before making a purchase will put themselves in the hands of psychotherapists with little or no prior investigation of their credentials, theoretical orientations, professional affiliations, or their records of successfully helping their clients in the past" (p77).

Relman (1998) referred to "stoned thinking" when the validity of therapy is based on "a melange of mystical intuition and emotional satisfaction" (Beyerstein 2001b).

Beyerstein (2001b) listed the warning signs for "fringe therapies" - its rationale is implausible in terms of scientific knowledge, there is a lack of supporting scientific evidence, or scientific studies have failed to support its claims. In relation to complementary and alternative medicine, but true for any therapy or treatment, Beyerstein (2001b) stated:

Clients should be dubious of any practitioner who (1) is ignorant of or hostile to mainstream science; (2) cannot supply a reasonable rationale for his or her methods; (3) uses promotional patter laced with allusions to spiritual forces and vital energies or to vague planes, vibrations, imbalances, and sensitivities; (4) claims to possess secret ingredients or processes; (5) appeals to ancient wisdom and "other ways of knowing"; (6) claims to "treat the whole person" rather than organ-specific diseases; or (7) claims to be persecuted by the establishment and encourages political action on his or her behalf, or is prone to attack or sue critics rather than responding with valid research (p236).

Beyerstein (2001b) lamented: "If only the ignorant and gullible were swayed by far-fetched claims, little else would be needed to explain the abundance of folly in modern society. But oddly enough, many people who are neither foolish nor ill-educated cling to beliefs repudiated by science" (p230). In response to this observation, he outlined three groups of explanations for the popularity of "fringe therapies":

1. Social and cultural reasons - including poor

scientific literacy, anti-scientific attitudes, and mistrust of traditional authority figures.

2. Psychological reasons - including wishful thinking, logical errors, and misattributions.

3. The apparent effectiveness of treatments - Many individuals improve after the use of these treatments, but it is due to another cause rather than the treatment itself (appendix 1A). These factors are true for any treatment (with or without a scientific basis).

1.1.1. Power Therapies

Deville (2005) used the term "power therapies" because of the claims of how powerful the therapies are (eg: one-session-cure). "It is claimed that these 'power therapies' are at the cutting-edge of psychiatry and psychology and that they are so termed because of their efficiency and efficacy being superior to traditional treatments" (Deville 2005 p438). Also called "alphabet therapies" because of the use of acronyms (Deville 2005).

Deville (2005) concentrated on six "power therapies":

a) Eye Movement Desensitisation and Reprocessing (EMDR (eg: Shapiro 1989) - Specific eye movements are used to aid PTSD sufferers.

b) Thought Field Therapy (TFT) (eg: Callahan and Callahan 1997) (appendix 1B) - A treatment for phobias that involves tapping fourteen "energy meridians" (or acupoints) on the surface of the skin in a specific order while thinking of the feared stimulus, and this realigns the disturbed emotions. Different orders of tapping are used for different fears.

c) Emotion Freedom Techniques (EFT) (eg: Craig 1999) - A variation of TFT which uses a single, general purpose, order of tapping for all fears.

d) Traumatic Incident Reduction (TIR) (eg: Gerbode 1989) - Discovering repressed chains of memories and recalling them under conditions of safety.

e) Tapas Acupuncture Technique (TAT) (eg: <http://www.tatlife.com/>) - Light pressure to certain areas of the body releases the blockage of energy that underlies psychological problems.

f) Visual-Kinaesthetic Dissociation (VKD) (eg: Bandler and Grinder 1979) - A variation of NLP where a trauma is re-imagined from different standpoints as if

watching a videotape.

Though each treatment has its own twist, they are loosely based on exposure and/or cognitive restructuring, which are effective strategies of cognitive-behavioural therapy. So it is possible that these "power therapies" will be better than no treatment, but, as Devilly (2005) noted, the "misattribution of treatment efficacy to accoutrement practices might lead me, for example, to hypothesise that it is the green leather chair that I use during imagined exposure for PTSD treatment which is the active ingredient to my therapy and responsible for symptom relief" (p441).

Yet many treatments like these have widespread followings, including from mental health professionals. Devilly (2005) described commonalities between the "power therapies" using Pratkanis's (1995) criteria of pseudoscience, which highlighted the social psychological processes involved:

i) Creation of a phantom - A currently unavailable goal that could be reached with a new approach.

ii) A rationalisation trap - A small series of steps that draw the individual into the treatment (often with increasing costs). "In effect, the target (eg: psychiatrist) rationalises that they must be interested as they have already invested substantial time and money into the practice" (Devilly 2005 p441).

iii) Source credibility and sincerity - The presence of a "guru-like leader" (ie: expert), with good intentions (ie: to help those suffering).

iv) "Granfalloon" ⁶ - Each treatment has its own self-regulation (ie: it is only possible to practice the therapy when a member of the official body and to become a member involves accepting the demands of the official body).

v) Self-generated persuasion - Those who have benefited from the treatment become its advocates. "It makes it very difficult for the decision maker (eg: psychiatrist or case manager) to appear to be a caring individual while publicly disbelieving the advocate or turning down approval for wholesale delivery of therapy X, having been told an anecdotal case study" (Devilly 2005 p442).

vi) Pre-persuasion - Setting criteria for "proof"

⁶ A term used by Kurt Vonnegut to describe a "proud and meaningless association of human beings" (Devilly 2005).

that suits the treatment rather than more objective criteria.

vii) Use of human heuristics - ie: cognitive short-cuts used in decision-making, like consensus bias (if lots of people agree, it must be true).

viii) Commonplaces - Commonly held beliefs in society, like science and medicine traditionally neglect the human spiritual side.

ix) Attack critics personally - eg: critics are uncaring. Devilly (2005) quoted a posting on a web forum for one of the "power therapies" - "Feeling rather tired of the scientism rampant in our professions, of the people who would rather devote their time to trying to disprove things than to helping people directly, and not feeling much patience right now - maybe I need to tap on that".

1.2. NEURO-LINGUISTIC PROGRAMMING

Neuro-Linguistic Programming (NLP) (eg: Bandler and Grinder 1979) is very popular today in many areas of life (eg: therapy, personal growth, education), and many claims about its success are made. It is based on the core assumption that individuals have internal representations of the world (the "map" or "preferred representational system"; PRS) that influences behaviour⁷. This is different to the actual world (the "territory"). Simplistically, the role of the NLP therapist is to help the individual with the gap between the map and the territory by various means (Wotkowski 2012)⁸.

There has been much debate about the claims and effectiveness of NLP. For example, Sharpley (1984) found no support "that NLP can help clients change" from fifteen studies. Einspruch and Forman (1985) countered that thirty-nine studies of NLP had methodological errors. Sharpley (1987) then reviewed these studies and five more to confirm no support for NLP. Heap (1988) made similar conclusions after analysing sixty-three studies⁹.

⁷ "The basic premise was that people are influenced by internal 'maps' of information which they gather and organise visually, aurally or kinaesthetically. It was claimed that the trained consultant could identify the method in which the information was stored by eye-gaze patterns, posture, tone of voice and language patterns. It was further claimed that this knowledge facilitated communication during therapy to effect change (eg: a kinaesthetic representational system would be more amenable to change through the use of 'feeling' words during therapy)" (Devilly 2005 p437).

⁸ There is no agreed definition of NLP (Sturt et al 2012).

⁹ "The conclusions drawn from a meta-analysis are only as sound as the studies that were included in it

Witkowski (2012) noted: "If the NLP claims concerning its instant effectiveness proved to be true, the field of psychotherapy would be transformed, perhaps even revolutionised, and research reporting the effectiveness of therapy would position NLP as a potential first-rate therapy. Nothing like this is taking place. Instead, experts warn against using NLP and classify it as one of the many dubious 'fringe therapies' (Beyerstein 2001a) or 'power therapies' (Deville 2005). NLP is also found on many lists of discredited therapies" (p31).

Norcross et al's (2006) experts gave NLP for mental and behavioural disorders a score of 3.87.

Admitting that the last review of NLP was over twenty years old, Witkowski (2012) performed a systematic review of empirical studies since Heap (1988). Ten studies (out of over 400 articles found) investigated the effectiveness of NLP, of which seven were non-supportive, two partially supportive, and one supportive. Witkowski (2012) concluded: "The present review is consistent with earlier reviews in finding that the vast majority of research studies have not supported either the fundamental tenets or the techniques of NLP. Among the 401 published papers on NLP, none were systematic research reviews supportive of the model" (p37).

Furthermore, he stated: "Despite the lack of scientific support, NLP has evolved into a big business, luring people with prospects of amazing changes, opportunities for great personal development, and uniquely effective psychotherapy results. This raises an important question: Is it ethical to promote an intervention that is devoid of scientific support" (Witkowski 2012 p37).

Sturt et al (2012) undertook a systematic review of experimental studies of NLP and health-related outcomes. Five randomised controlled trials and five pre-post studies of relevance were found. There was great variety between the studies in, for example, health condition studied (eg: anxiety disorders, morning sickness), participants (eg: more female; age range from 17-75 years old), and amount of NLP (between 3-20 1-2-hour sessions). Most outcome measures were self-reported, and risk of bias and incomplete data were high. Overall, there was "little evidence" that NLP was beneficial for the health conditions studied.

and the judgment of the reviewer who chooses and weights them" (Beyerstein 2001a p72).

1.3. APPENDIX 1A - APPARENTLY GETTING BETTER

There are a number of reasons why an individual could get better either without treatment or even if the treatment is ineffective. A key element is the difference between "disease" (the physiological aspect) and "illness" (the subjective feelings). The subjective response to symptoms is influenced by attitudes, expectations, and other psychological processes (Beyerstein 2001b).

1. The disease runs its natural course - Many diseases have a time-limited course, and thus individuals will improve anyway after this (with or without treatment).

2. Spontaneous remission - This is where a disease takes an unexpected upturn which seems to have no clear explanation. Beyerstein (2001b) quoted an oncologist who had seen twelve cases of cancerous tumours disappearing without treatment out of 6000 patients treated.

3. Some diseases are cyclical - Diseases like arthritis and asthma have ups and downs in severity. Treatment given during the downturn can appear to work when it is just an upturn.

4. Not sick in the first place - There is the term "worried well" to describe individuals who focus on minor symptoms as a sign of serious disease. The extreme version of this is hypochondria. Such individuals may be helped (psychologically) by a treatment that has no physiological benefits.

5. The benefits of the "sick role" - Some individuals may gain from taking on the role of a sick person, and they do not respond well to traditional medicine telling them that they are well. Such individuals then seek alternative treatments, which give them the attention they desire, for example, and appear to "cure" them.

6. Misdiagnosis - Individuals who have been misdiagnosed with a disease, then recover with a particular treatment. The credit is given to the treatment for the success.

7. Spin-offs from the treatment - For example, individuals may improve with an alternative treatment that gives them attention (which traditional medicine does not) rather than the "active ingredient" of the treatment. In one sense, this is a placebo effect.

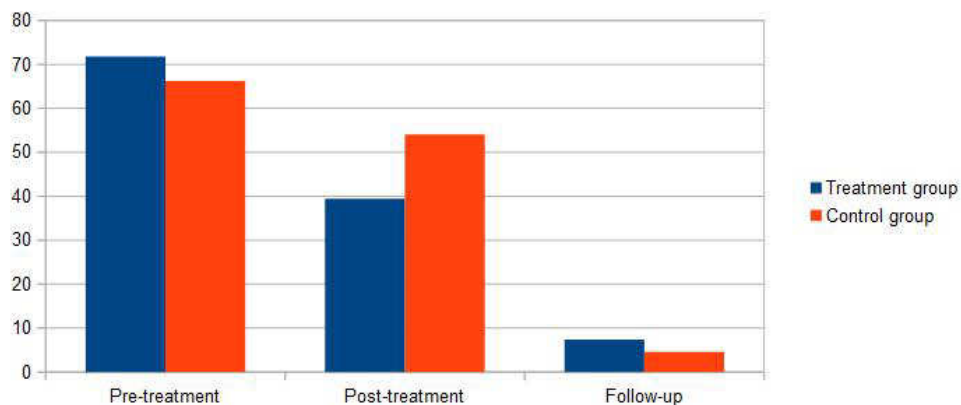
8. The placebo effect - "Through a combination of

suggestion, expectancy, and cognitive reinterpretation, patients given biologically useless treatments often can experience subjective relief" (Beyerstein 2001b p234).

1.4. APPENDIX 1B - THOUGHT FIELD THERAPY (TFT)

Connolly and Sakai (2011) reported the use of TFT with 145 adult survivors of genocide in Rwanda in 1994, who showed symptoms of PTSD. The participants were randomised to receive treatment (n = 71) or to a waiting-list control group (n = 74). Baseline measures of symptoms were taken using standardised measures, like the Modified PTSD Symptom Scale (MPSS) (Falsetti et al 1993)¹⁰.

The treatment group showed significant decreases in symptoms between baseline and seven days after treatment and at two-year follow-up (figure 1.1). Connolly and Sakai (2011) confidently stated: "Results suggest that TFT significantly reduces symptoms of anxiety, depression, anger/irritability, defensive avoidance, dissociation, impaired self-reference, and tension reduction behaviours, as well as the severity and frequency of PTSD symptoms."



(Data from Connolly and Sakai 2011 table 3 p170)

Figure 1.1 - Percentage of participants rated as PTSD using MPSS criterion.

Table 1.1 lists some of the key limitations of the study and how they compare to the ideal. Many of them are common to studies of any type of treatment.

¹⁰ Seventeen symptoms of PTSD are self-reported for frequency (0-3) and severity (A-E).

Connolly and Sakai (2011)	Ideal Study
1. Comparison with waiting list.	1. Comparison with standard treatment, placebo, or other treatment.
2. Self-reported measures of symptoms.	2. Diagnosis by third party (eg: psychiatrist), but standardised self-reported measures are common.
3. Volunteers.	3. Random sample, but practicalities and/or ethics lead to the use of volunteers.
4. Treatment performed by 28 trained local therapists.	4. Same therapist throughout, but unlikely with large-scale studies.
5. Sample size.	5. Larger is assumed to be better, but practicalities can limit the size.
6. Use and translation of Western psychometric measures to local language.	6. There is an issue that any measure designed in one country or area of the world will be inherently biased.

Table 1.1 - The key limitations of Connolly and Sakai (2011) as compared to an ideal study.

1.5. REFERENCES

Bandler, R & Grinder, J (1979) The Structure of Magic: A Book About Language and Therapy Palo Alto, CA: Science and Behavior Books

Beyerstein, B.L (2001a) Fringe psychotherapies: The public at risk The Scientific Review of Alternative Medicine 5, 70-79

Beyerstein, B.L (2001b) Alternative medicine and common errors of reasoning Academic Medicine 76, 3, 230-237

Callahan, R.J & Callahan, J (1997) Thought field therapy: Aiding the bereavement process. In Figley, C.R & Bride, B.E (eds) Death and Trauma: The Traumatology of Grieving Philadelphia: Taylor & Francis

Connolly, S & Sakai, C (2011) Brief trauma intervention with Rwandan genocide-survivors using thought field therapy International Journal of Emergency Mental Health 13, 3, 161-172

Craig, G (1999) Emotional Freedom Techniques (revised ed) Sea Ranch, CA: Author

Dawes, R.M (1994) House of Cards: Psychology and Psychotherapy Built on Myth New York: Free Press

Deville, G.J (2005) Power therapies and possible threats to the science of psychology and psychiatry Australian and New Zealand Journal of Psychiatry 39, 437-445

Einspruch, E.L & Forman, B.D (1985) Observations concerning research literature on Neuro-linguistic Programming Journal of Counseling Psychology 32, 589-596

Falsetti, S.A et al (1993) The modified PTSD symptom scale: A brief self-report measure of Post-Traumatic Stress Disorder The Behaviour Therapist 16, 161-180

- Gerbode, F (1989) Beyond Psychology: An Introduction to Metapsychology (2nd ed) Palo Alto, CA: IRM Press
- Heap, M (1988) Neuro-linguistic programming: An interim verdict. In Heap, M (ed) Hypnosis: Current Clinical, Experimental and Forensic Practices London: Croom Helm
- McQueen, D.V (2001) Strengthening the evidence base for health promotion Health Promotion International 16, 3, 261-268
- Norcross, J.C et al (2006) Discredited psychology treatments and tests: A Delphi poll Professional Psychology: Research and Practice 37, 515-522
- Pratkanis, A.R (1995) How to sell a pseudoscience Skeptical Inquirer 19, 19-25
- Relman, A (1998) A trip to Stonesville The New Republic 378, 28-37
- Rosen, R.D (1978) Psychobabble: Fast Talking in the Era of Feeling New York: Atheneum
- Shapiro, F (1989) Efficacy of Eye Movement Desensitisation procedure in the treatment of traumatic memories Journal of Traumatic Stress 2, 199-223
- Sharpley, C.F (1984) Predicate matching in NLP: A review of research on the preferred representational system Journal of Counseling Psychology 31, 238-248
- Sharpley, C.F (1987) Research findings on Neuro-linguistic Programming: Non-supportive data or untestable theory? Journal of Counseling Psychology 34, 103-107
- Sturt, J et al (2012) Neurolinguistic programming: A systematic review of the effects on health outcomes British Journal of General Practice 62, 604, e757-e764
- Witkowski, T (2012) A review of research findings on neuro-linguistic programming The Scientific Review of Mental Health Practice 9, 1, 29-40

2. REPLICATION AND THE "MANY LABS" PROJECT

Klein et al (2014) began: "Replication is a central tenet of science; its purpose is to confirm the accuracy of empirical findings, clarify the conditions under which an effect can be observed, and estimate the true effect size... Successful replication of an experiment requires the recreation of the essential conditions of the initial experiment. This is often easier said than done" (p142). Failure to replicate findings can mean that results are inappropriately generalised (eg: from one culture to all cultures; from the laboratory to real-life), or even that "spurious findings" remain unchallenged.

The "Many Labs" Replication Project ¹¹ ¹² set out to replicate thirteen effects/studies in psychology across thirty-six samples and settings. In total, data were collected from 6344 participants (from 27 laboratory settings and nine online; 25 from the USA and eleven from elsewhere).

The thirteen effects/studies were:

1. Sunk costs.

Oppenheimer et al (2009) asked 113 New York University students to imagine a scenario where they had a ticket to see their favourite football team play an important game, but the weather is very unpleasant to spectate (ie: cold). The participants rated how likely they were to go either if the ticket was free or they had paid for it. They were more likely to go in the latter condition (mean: 7.46 vs 6.93 for free ticket, out of nine, where 9 = "definitely go to the game").

2. Gain versus loss framing.

Tversky and Kahneman (1981) offered choices of a certain win versus a gamble (positive framing), or a certain loss versus a gamble (negative framing). For example, a win of 20 points or a gamble with a one-third probability of 60 points and a two-thirds probability of nothing. For gains, participants preferred the certain choice, but the opposite for losses (table 2.1).

3. Anchoring.

Giving participants an anchor (ie: an example)

¹¹ Information and data at <https://osf.io/ydpbf/>.

¹² See also the Reproducibility Project at <https://osf.io/ezcu/wiki/home/>.

Tversky and Kahneman (1981) offered the following choices to students ¹³ :

Problem 1

Imagine that there is an outbreak of a rare disease, which is expected to kill 600 people. Two alternative programmes to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programme are as follows:

- If programme A is adopted, 200 people will be saved.
- If programme B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Tversky and Kahneman found that 72% chose A and 28% B. This is because when positive framing is used ("saved") individuals tend to be less risky.

Problem 2

Again the same scenario, but with these options:

- If programme C is adopted 400 people will die.
- If programme D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Tversky and Kahneman found that 22% choose C and 78% programme D. When it is negatively framed ("deaths") individuals tend to be riskier.

Table 2.1 - Two problems used by Tversky and Kahneman (1981).

influences their subsequent estimation of size (eg: height of Mount Everest; table 2.2), even if the anchor is clearly too large or small (Jacowitz and Kahneman 1995).

Anchor	Median estimate
None	12 000
2000	8000
45 500	42550

(Data from Jacowitz and Kahneman 1995 table 1 p1163)

Table 2.2 - Estimated height of Mount Everest (in feet) based on anchors used ¹⁴.

¹³ In each case the same outcome is being offered though the wording is different - ie: 400 people die and 200 don't.

¹⁴ Actual height = around 30 000 feet.

4. Retrospective gambler's fallacy ¹⁵.

Oppenheimer and Monin (2009) told participants about a man rolling dice in a casino. The gambler either got three 6s from the three dice, two 6s from two dice, or two 6s and a 3 from three dice. Then the participants were asked how many times the man had rolled the dice before that throw. Participants who had heard about the three 6s gave a significantly larger estimate than the other two conditions (figure 2.1).

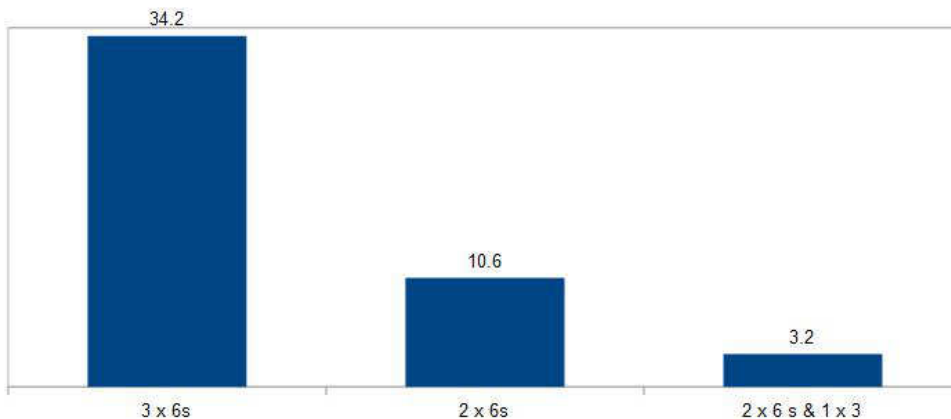


Figure 2.1 - Mean estimate of number of rolls of dice before throw in story.

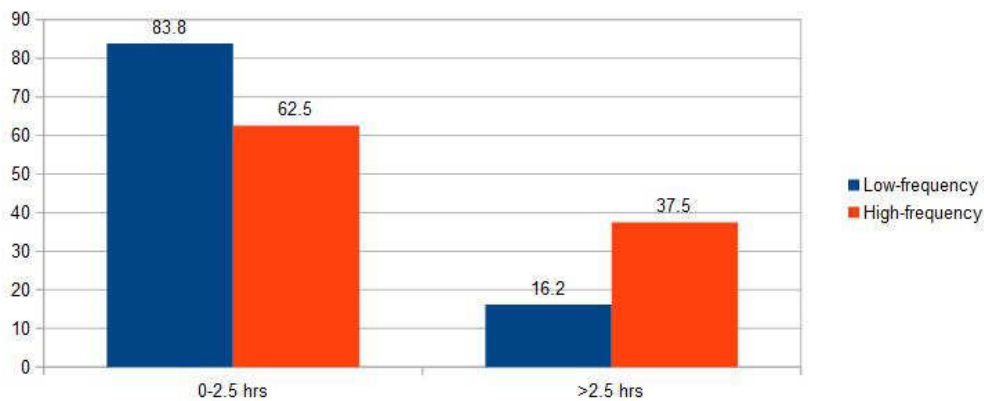
5. Low-versus-high category scales.

Schwarz et al (1985) showed that the response options offered influenced the answers to questions about frequency. For example, participants were asked how much television they watched per day. In the low-frequency condition, the six choices ranged from "up to half an hour" to "more than two and a half hours", whereas in the high-frequency condition, the range was "up to two and a half hours" to "more than four and a half hours". The average number of hours was lower in the low-frequency condition (2.7 hours vs 3.2 hours in high-frequency condition) (figure 2.2).

6. Norm of reciprocity.

Hyman and Schwarz (1950) found that a decision about allowing the behaviour of another group is influenced by the perception of allowing your group. Using an example

¹⁵ This is defined as "an event that seems rare appears to come from a longer sequence than an event that seems more common" (Oppenheimer and Monin 2009 p326).



(Data from Schwarz et al 1985 table 1 p391)

Figure 2.2 - Percentage of participants based on response options given.

of the time, the researchers asked whether US reporters should be allowed into communist countries, and whether communist reporters should be allowed into the USA. Participants answered "yes" to the latter question more often when the questions were asked in this order than the other way around.

7. Allowed/forbidden wording.

Question phrasing influences the response. Rugg (1941) asked the same question in two ways, for example - should speeches against democracy be allowed or should speeches be forbidden? Over 60% of participants said "no" to the allowed version, and less than 50% said "yes" to the forbidden version.

8. Source attribution.

The source of a piece of information influences its acceptance. Lorge and Curtiss (1936) presented the same quotation to US participants as from a famous American or a famous Russian. More participants agreed with the sentiment of the quote when the source was the American.

9. Flag priming influencing conservatism.

Carter et al (2011) presented US participants with photographs of the American flag or not in an apparently unconnected task before measuring their political attitudes. Participants who had seen the flag had

significantly more conservative attitudes.

10. Currency priming influencing system justification.

Caruso et al (2013) found that participants primed to think about money supported the current social system in an attitude questionnaire more than individuals not primed (mean: 4.96 vs 3.99 out of 7). The priming was done by asking the questions on paper with a faint watermark of dollar bills or plain paper.

11. Imagined contact reducing prejudice.

Husnu and Crisp (2010) asked British non-Muslim participants to imagine meeting and talking to a British Muslim stranger or imagine walking alone. The former group reported more positive attitudes towards British Muslims afterwards.

12. Sex differences in implicit attitudes towards mathematics.

Nosek et al (2002) found more negative implicit attitudes among women at Yale University towards mathematics than arts as compared to men.

13. Implicit attitudes towards mathematics relates to self-reported (explicit) attitudes towards mathematics.

Nosek et al (2002) also found that explicit attitudes about mathematics correlated with the implicit attitudes.

Overall, Klein et al (2014) found that ten effects were replicated as in the original study (though the size of the effect varied). "Imagined contact reducing prejudice" had weak replication (eg: 4 of 36 samples and one contradiction), while "flag priming influencing conservatism" and "currency priming influencing system justification" were not replicated.

The researchers were positive about the "Many Labs" project which used crowdsourcing and open data: "It allows for increased confidence in the existence of an effect and for the investigation of an effect's dependence on the particular circumstances of data collection... Further, a consortium of laboratories could provide mutual support for each other by conducting similar large-scale investigations on original

research questions, not just replications. Thus, collective effort could accelerate the identification and verification of extant and novel psychological effects" (Klein et al 2014 p151).

But the replications were not identical to the original studies, particularly as aspects of the methodology were changed to fit the situation today. For example, Hyman and Sheatsley (1950) were working during the Cold War, so the use of US and communist reporters was relevant. Klein et al (2014) changed communist to North Korea to ensure "a suitable modern target". Similarly, Lorge and Curtiss (1936) had used Thomas Jefferson (liked individual) and Vladimir Lenin (disliked individual), which were changed to George Washington (liked person) and Osama Bin Laden (disliked person).

REFERENCES

Carter, T.J et al (2011) A single exposure to the American flag shifts support toward Republicanism up to eight months later Psychological Science 22, 1011-1018

Caruso, E.M et al (2013) Mere exposure to money increases endorsement of free-market systems and social inequality Journal of Experimental Psychology: General 142, 301-306

Husnu, S & Crisp, R.J (2010) Elaboration enhances the imagined contact effect Journal of Experimental Social Psychology 46, 943-950

Hyman, H.H & Sheatsley, P.B (1950) The current status of American public opinion. In Payne, J.C (ed) The Teaching of Contemporary Affairs: 21st Yearbook of the National Council of Social Studies New York: National Council of Social Studies

Jacowitz, K.E & Kahneman, D (1995) Measures of anchoring in estimation tasks Personality and Social Psychology Bulletin 21, 1161-1166

Klein, R.A et al (2014) Investigating variation in replicability Social Psychology 45, 3, 142-152

Lorge, I & Curtiss, C.C (1936) Prestige, suggestion, and attitudes The Journal of Social Psychology 7, 386-402

Nosek, B.A et al (2002) Math = male, Me = female, therefore ≠ Me Journal of Personality and Social Psychology 83, 44-59

Oppenheimer, D.M & Monin, B (2009) The retrospective gambler's fallacy: Unlikely events, constructing the past, and multiple universes Judgment and Decision-Making 4, 326-334

Oppenheimer, D.M et al (2009) Instructional manipulation checks: Detecting satisficing to increase statistical power Journal of Experimental Social Psychology 45, 867-872

Rugg, D (1941) Experiments in wording questions: II Public Opinion Quarterly 5, 91-92

Schwarz, N et al (1985) Response scales: Effects of category range on reported behaviour and competitive judgments Public Opinion Quarterly 49, 388-395

Tversky, A & Kahneman, D (1981) The framing of decisions and the psychology of choice Science 211, 453-458