

Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice

VOLUME 46

2017

NUMBER 1

Contents

- | | |
|--|----|
| A Not so Bogus Pipeline: A Study of the Bogus Pipeline Effect and Its Implications for Polygraph Testing
Elizabeth Elliott, Vincent Egan and Don Grubin | 1 |
| The Effect of Questions on Suspects' Perception of Evidence in Investigative Interviews: What Can We Infer from the Basic Literature?
Laure Brimbal, Maria Hartwig, and Angela M. Crossman | 10 |
| Assessing Polygraph Accuracy; The Importance of Choosing an Evaluation Technique Which Is Compatible with the Way the Examinations Were Conducted
Avital Ginton | 40 |
| The Use of Polygraph Testing for Theft Investigation in Private Sector Institutions
Laura Pérez-Campos Mayoral, Eduardo Pérez-Campos Mayoral
Gabriel Mayoral Andrade, Carlos Perezcampos Mayoral,
Rocío Martínez Helmes, Eduardo Pérez-Campos | 44 |
| Bayesian Probabilities of Deception and Truth-telling for Single and Repeated Polygraph Examinations
Raymond Nelson and Finley Turner | 53 |

Polygraph

Editor-in-Chief: Mark Handler
E-mail: Editor@polygraph.org
Managing Editor: Nayeli Hernandez
E-mail: polygraph.managing.editor@gmail.com

Associate Editors: Réjean Belley, Ben Blalock, Tyler Blondi, John Galianos, Don Grubin, Maria Hartwig, Charles Honts, Matt Hicks, Scott Hoffman, Don Krapohl, Thomas Kuczek, Mike Lynch, Ray Nelson, Adam Park, David Raskin, and Stuart Senter.

APA Officers for 2016 – 2017

President – J. Patrick O'Burke
E-mail: president@polygraph.org

President Elect – Jamie McCloughan
E-mail: presidentelect@polygraph.org

Chairman – Walt Goodson
E-mail: chairman@polygraph.org

Director 1 – Barry Cushman
109 Middle Street
Portland, ME 04101
E-mail: directorcushman@polygraph.org

Director 2 – Raymond Nelson
E-mail: directornelson@polygraph.org

Director 3 – George Baranowski
1912 E US Hwy 20, Suite 202
Michigan City, IN 46340
E-mail: directorbaranowski@polygraph.org

Director 4 – Sabino Martinez
E-mail: directormartinez@polygraph.org

Director 5 – Steve Duncan
E-mail: directorduncan@polygraph.org

Director 6 – Darryl Starks
E-mail: directorstarks@polygraph.org

Director 7 – Gary F. Davis
E-mail: directordavis@polygraph.org

Director 8 – Dan Violette
E-mail: directorviolette@polygraph.org

Treasurer – Chad Russell
E-mail: treasurer@polygraph.org

General Counsel – Gordon L. Vaughan
E-mail: generalcounsel@polygraph.org

Seminar Chair – Michael Gougler
E-mail: seminarchair@polygraph.org

National Officer Manager – Lisa Jacocks
Phone: 800-APA-8037; (423)892-3992
E-mail: manager@polygraph.org

Assistant Office Manager – Stephanie Prairie
E-mail: apaoffice@polygraph.org

Subscription information: *Polygraph* is published semi-annually by the American Polygraph Association. Editorial Address is Editor@polygraph.org. Subscription rates for 2017: One year \$150.00 (Domestic), \$180.00 (Foreign). Change of address: APA National Office, P.O. Box 8037 Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN *POLYGRAPH* DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

Instructions to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript Submission

Manuscripts must be in English, and may be submitted, along with a cover letter, on electronic media (MS Word). The cover letter should include a telephone number, and e-mail address. All manuscripts will be subject to a formal peer-review. Authors may submit their manuscripts as an e-mail attachment with the cover letter included in the body of the e-mail to:

Editor@polygraph.org

As a condition of publication, authors agree that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere. Authors also agree to give reasonable access to their data to APA members upon written request.

Manuscript Organization and Style

All manuscripts must be complete, balanced, and accurate. Authors should follow guidelines in the *Publications Manual of the American Psychological Association*. The manual can be found in most public and university libraries, or it can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Writers may

exercise some freedom of style, but they will be held to a standard of clarity, organization, and accuracy. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psychophysiological Detection of Deception* (2012) which is available from the national office of the American Polygraph Association. Legal case citations should follow the *West* system.

Manuscript Review

An Associate Editor will handle papers, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 4 weeks of receipt of manuscript. Articles submitted for publication are evaluated according to several criteria including significance of the contribution to the polygraph field, clarity, accuracy, and consistency.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be granted if the senior author approves.

Greetings from the Editor's Desk

As you probably noticed, the name of the APA peer-reviewed journal has been changed to Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice.

The board unanimously voted this change to reflect the increased scope of the journal. Our goal is to seek credibility assessment contributions beyond those related solely to polygraph. As credibility assessment professionals, our goal should be to assist end-users and consumers of polygraph in their decision-making processes. To that end, our board felt we should expand the journal content, and I heartily agree.

We hope you find the expanded scope useful and educational. As always we are humbled by the trust placed in us.

Sincerely,

Mark Handler Editor and Nayeli Hernandez Managing Editor.

A Not So Bogus Pipeline: A Study of the Bogus Pipeline Effect And Its Implications for Polygraph Testing

Elizabeth Elliott, Vincent Egan and Don Grubin

The University of Nottingham

School of Medicine

Division of Psychiatry and Applied Psychology

Nottingham, United Kingdom

Abstract

Purpose

The Bogus Pipeline effect refers to a procedure in which individuals are made to believe that their responses to questions will be independently verified by an infallible lie detector, resulting in an associated reduction in socially desirable responses. Polygraph testing is known to elicit increased disclosure by examinees, which some critics claim is an example of this effect. We explored whether a bogus lie detector said to have just 75% accuracy (a level below that of polygraph testing) would be as effective in eliciting disclosures as a 100% accurate lie detector when participants were questioned about cheating in a competition.

Method

Participants took part in tasks as part of a group, with the winning group receiving £500 to donate to a charity. Each group contained a confederate who cheated on one of the tasks. Participant later returned for a debrief, during which they were asked about cheating within the group. Each was randomized to one of three conditions: one in which they were questioned with the aid of an apparatus they were told was a nearly 100% accurate lie detector, one with a 75% accurate 'lie detector' was used, and a control condition.

Results

A significantly larger proportion of participants in both bogus lie detector conditions than in the control condition disclosed cheating by a group member, but there was no significant difference between the two bogus pipeline groups.

Conclusion

The bogus pipeline effect may not be bogus after all.

Keywords: Bogus pipeline; disclosure; lie detection

Introduction

Proponents of polygraph testing often maintain that one of its most beneficial aspects of the paradigm is the way in which it facilitates the disclosure of large amounts of information that would otherwise remain concealed (English et al, 2003; Grubin, 2008).

In some settings such as pre-employment screening and sex offender management, this 'utility' aspect of the procedure is sometimes said to be as important as its accuracy. Critics, however, argue that the effect may simply reflect the examinee's belief that the polygraph 'works' and would disappear if this view was not held (Saxe, 1991; Cross & Saxe, 2001;



Rosky, 2013). This is typically referred to as 'the bogus pipeline' (BPL) effect. Though the BPL effect has been demonstrated many times, neither its strength nor its predictability is as strong as often assumed (Roese & Jamieson, 1993; Aguinis et al., 1995), and its actual relevance to polygraph testing is largely unquantified.

The 'bogus pipeline' was first described and titled by Jones and Sigall (1971), who referred to it as "a direct pipeline to the soul (or some nearby location)" (pp. 349). They outlined a technique in which subjects were made to believe that their responses to questions will be independently verified by some form of device that was an infallible lie detector, with an associated reduction in socially desirable responses. The procedure has been applied to a wide range of attitudes and behaviours such as racist beliefs (Sigall & Page, 1971), interpersonal attraction (Page & Moss, 1975), cheating (Fisher & Brunell, 2014), sexual activity (Alexander & Fisher, 2003), and cognitive distortions in sex offenders (Gannon et al., 2007), in general producing what is considered to be more truthful self-report: a review by Roese & Jamieson (1993) found a shift towards more 'negative' self-report under BPL conditions in about two thirds of studies, while their associated meta-analysis reported a moderate effect size of $d=.41$ (confidence interval .33 - .49).

The actual form the BPL takes varies between studies, but in general it is composed of three elements. First, participants are introduced to or told about a device that can measure their attitudes or detect lies (the device, of course, does not measure anything). Participants are then either attached to the apparatus with a few rigged demonstrations taking place to convince them of its accuracy, or the procedure by which their responses will be checked is explained. In the final step participants are instructed to answer questions as honestly as possible either while attached to the device or in anticipation of their answers being checked by it, although in some studies they are instead asked to predict the device's findings. Clearly, participants need to believe what they have been told about the device's lie detection capabilities, but most studies find this is usually the case.

The BPL effect appears to depend on the beliefs or behaviours in question being perceived by the individual as socially undesirable (Roese & Jamieson, 1993; Aguinis et al., 1995). Its mode of action, however, is uncertain. It may reflect what has been referred to as 'self-preservation', that is, avoidance of being considered a liar or being out of touch with one's true beliefs, although it has also been suggested that it may simply encourage people to attend more closely to how they are answering questions.

There is some evidence that gender differences in disclosure diminish when participants believe their responses are being monitored by a lie detector (Fisher & Brunell, 2014). It is possible that personality moderates the impact of the BPL, however this has not been previously researched.

As indicated above, studies of the BPL effect make use of a putative 100% lie detector on the assumption that it depends on a participant's belief that the lie detecting machine is virtually infallible, with the implication being that the effect would be lost if this is not the case. In reality, of course, no such device exists, which makes actual methods of detecting deception vulnerable if one of their goals is to encourage disclosures. This assumption, however, has never been tested.

If a 100% 'lie detector' is not in fact necessary for the BPL effect, and enhanced disclosure is elicited at levels of perceived accuracy less than this, then claims that procedures such as polygraph testing derive their efficacy in eliciting disclosures by making false claims about their accuracy lose their force (provided, of course, that their accuracy levels are greater than that required to elicit the BPL effect). The aim of the current study, therefore, is to evaluate the effect on disclosures when the 'pipeline' is not presented as being infallible. The influence of individual differences in personality and suggestibility upon levels of disclosure are also explored.



Method

Participants

Participants were undergraduate students at an English University who were studying a range of disciplines. They were offered £20 to take part in the study provided that they attended each of two sessions.

The aim was to recruit 155 participants. This was determined following a power calculation based on an expected effect size of $d=.41$ (Roese & Jamieson, 1993), $p < .05$, with 80% power (G*Power: Erdfelder, Faul, & Buchner, 1996). In the event, 180 participants signed up to take part in the study of whom 145 attended both parts 1 and 2, but data was complete for only 141 (84 females and 57 males). IQ wasn't measured, but as all were undergraduate students it was assumed they were of roughly similar intellectual ability.

Procedure

Prior to the study, participants were briefed in large groups using a standard study description. They were told that they were assisting with research exploring how and whether group performance impacts on an individual's processing speed on subsequent tasks. Oral consent was obtained, and participants were informed that they could withdraw from the research at any time.

Participants were allocated into groups of 5-10, each of which included a confederate who was a fellow university student. They were informed that a £500 donation to the charity of their choice would be awarded to the group with the highest overall score on the tasks they would be completing.

Stage 1

While in their groups, participants completed a brief psychometric test battery (the components of which are described below). They were then asked to undertake three tasks, two on an individual basis and one as a group:

- each participant completed an old version of the Wechsler's (WAIS) digit symbol task in 120 seconds;

- this was followed by a 5 minute group task in the form of a multiple choice quiz;
- in the final stage participants were asked to complete an anagram task individually, but while sat in their group. This was the 'Words Task' (Wiltermuth, 2011), which is an objective assessment of cheating – it must be completed in order, but the third anagram is in fact impossible to solve, which means that no one should be able to proceed beyond it.

At the completion of the tasks the researcher asked one member of the group to add up the individual scores to obtain the total for the group. The researcher left the room after this instruction, making an excuse for doing so. While the researcher was out of the room participants submitted their scores to the group member responsible for compiling the overall score. When it was his (or her) turn, the confederate informed the other group members that they had cheated by skipping number three on the anagram task, thereby increasing the group total score and making the others aware that one of their colleagues had cheated.

Stage 2

Participants were asked to return for the second part of the study approximately four weeks later ostensibly to answer a number of questions about the testing experience. The same researcher interviewed all participants.

In Stage 2 individuals were randomly allocated to one of three conditions using a random number generator: one in which they were questioned with the assistance of a 100% accurate 'lie detector', one using a 'lie detector' that was 75% accurate, and a control condition in which they were questioned without any external apparatus (the former two are referred to as the BPL groups). 42 participants (30%) were allocated to the control condition, 49 (35%) to the 75% condition, and 50 (36%) to the 100% condition. Participants were not asked how accurate they believed the 'lie detector' to be prior to this, as the study was



more focused on testing proof of concept, and asking such questions may have primed and encouraged them to challenge information given by the researcher.

Participants were informed that some groups reported abnormally high scores suggestive of dishonesty amongst some of those taking part. The researcher explained that it was important to determine which group was dishonest given the sizeable charity donation that was at stake. Individuals were asked whether they or anyone in their group had cheated (as a combined single question). The presumption is that individuals would be more likely to admit to group cheating if they believe the lie detector would otherwise show them to be deceptive. This is due to the influence of the BPL in circumventing social desirability biases, and encouraging disclosure.

In the two BPL groups the lie detector was introduced as a means of determining whether the group result should be discounted. Participants in these groups were informed of the machine's accuracy in detecting deception by measuring vital signs such as heart rate, breathing and galvanic skin response; they were told either that it was nearly 100% accurate, or accurate 75% of the time, depending on the group they were in. The assumption of being in the 75% group, is that if participants know that the machine is not 100% effective, they always have to opportunity of appearing to being an exception to the phenomena, thus undermining its perceived effectiveness. Components from a genuine polygraph instrument were attached to participants, and to enhance the believability of the bogus device they were instructed to reply 'yes' to two questions, one which evoked a lie response ('Is today Saturday?'), and one a true response ('Are we in Durham, United Kingdom?'). They were then shown a fabricated screenshot that purported to differentiate the false from the truthful response.

Participants in all three groups were asked the following:

'Did you yourself or anyone in your group attempt to cheat at any point of the experiment?'

Because there was a confederate cheating in all conditions, there was an opportunity for all participants to be deceptive. Participants in the control group were given the opportunity to be deceptive in the absence of a lie detector machine. For the two experimental conditions, those who did not disclose were informed that the lie detecting machine had indicated deception. The question was then repeated with the researcher reminding the participant to be as truthful as possible, 'it is important that you answer as honestly as possible,' this offered another opportunity for disclosure. Those who disclosed cheating were not asked any further questions.

At the conclusion of the questioning all participants were debriefed and informed of the true nature of the study. They were asked not to discuss the manipulation with others until the study was completed.

Measures

Personality

The NEO International Personality Item Pool (short version) (IPIP-50; Goldberg, 1999) is a questionnaire comprised of 50 items registering five dimensions: extraversion, agreeableness, conscientiousness, emotional stability and intellect. Each domain comprises ten statements (scored 1–5). Average alpha values for trait dimensions are approximately 0.80. Valid IPIP-50 profiles were obtained for 141 participants.

Suggestibility

The Short Suggestibility Scale (SSS) from the Multidimensional Iowa Suggestibility Scale (MISS; Kotov, Bellman, & Watson, 2007) was used. It consists of 21 items each scored 1–5 based on the extent of agreement with each statement ('not at all' to 'a lot'). Higher scores are indicative of increased levels of suggestibility. This measure was used to explore whether individuals with higher levels of suggestibility were more or less likely to disclose in any of the three conditions.



Ethics

Ethical approval was granted by the University of Nottingham, Faculty of Medicine & Health Sciences Research Ethics Committee.

Results

The Bogus Pipeline Effect

Overall, 67% of respondents disclosed cheating behaviour (Table I). In the 75% condition all of the disclosures were made in the

pre-test stage of the procedure; none of the 12 who were told the 'failed' the test went on to disclose. In the 100% condition, 41 of the 43 disclosures were in the pre-test stage, while just 2 of the 9 individuals who originally denied that cheating took place disclosed *after* being informed they were found to be deceptive on the lie detector machine.

A chi-square test of independence was performed to examine the relationship between bogus pipeline condition and disclosure. The relationship between these variables was significant ($\chi^2 = 28.52$; $p < .0001$) with a medium

Table 1. Number and per percentage in each group who disclosed cheating behaviour

	No Disclosure	Disclosure
Control	27 (64%)	15 (36%)
75% lie detector	12 (25%)	37 (76%)
100% lie detector	7 (14%)	43 (86%)
Total	46	95

to large effect size of $V = .45$ (Cohen, 1988).

Individuals in the 100% condition were no more likely to disclose than those in the 75% condition ($\chi^2 = 1.76$, $p = .18$). However, those in the 100% and 75% conditions were each significantly more likely to disclose than those in the control condition ($\chi^2 = 14.63$; $p = .0001$; $V = 0.40$ and $\chi^2 = 24.77$, $p < .0001$; $V = 0.52$ respectively).

Gender

No relationship was found between gender and disclosure between the three groups ($\chi^2 = .26$, $p = .61$), or between the two experimental groups ($\chi^2 = .47$, $p = .49$).

Personality

One hundred thirty six participants completed both the IPIP-50 (Goldberg, 1992) and both stages of the study. A one way between subjects ANOVA found no significant relationship between any of the personality factors and disclosure (extroversion $F(1, 134) = .37$, $p = .54$; conscientiousness $F(1, 134) = .27$, $p = .60$; emotional stability $F(1, 134) = .87$, $p = .36$; agreeableness $F(1, 134) = 1.65$, $p = .60$; intellect $F(1, 134) = .87$, $p = .20$).

A series of separate independent two-way ANOVA's were conducted to explore the impact of personality traits on disclosure between the two BPL conditions. Again, no significant differences were found: extroversion



main effect $F(1, 27) = 1.0$, $p = .44$ and interaction between extroversion and disclosure $F(1, 18) = 1.1$, $p = .41$; agreeableness main effect $F(1, 23) = 1.5$, $p = .13$ and interaction between agreeableness and disclosure $F(1, 19) = .60$, $p = .89$; conscientiousness main effect $F(1, 27) = .63$, $p = .90$ and interaction between conscientiousness and disclosure $F(1, 18) = .83$, $p = .66$; emotional stability main effect of $F(1, 30) = .82$, $p = .72$ and interaction between emotional stability and disclosure $F(1, 18) = .84$, $p = .62$; intellect main effect $F(1, 24) = 2.17$, $p = .26$ and interaction between intellect and disclosure $F(1, 18) = .84$, $p = .34$.

Suggestibility

One hundred thirty-three participants completed the SSS measure of suggestibility. An independent-samples *t*-test was used to compare levels of suggestibility between those who disclosed and those who did not. No significant difference in the scores was found: $M = 52.86$, $SD = 10.03$ for those who disclosed, $M = 51.93$, $SD = 9.22$ for those who did not disclose; $t = .52$, $p = .60$. A separate independent *t*-test comparing suggestibility scores between the 75% BPL condition ($M = 51.44$, $SD = 9.27$) and the 100% ($M = 51.06$, $SD = 12.82$) BPL condition was also not significant: $t = .16$, $p = .34$.

Discussion

Participants in both Bogus Pipeline conditions were more likely to disclose cheating behaviour than those in the control group with medium to large effect sizes, but there was no difference between the two bogus pipeline conditions themselves – whether participants were informed that the lie detecting machine was 100% accurate or accurate only 75% of the time did not have a material impact on the proportion of those who disclosed. Thus, although it is generally assumed that the BPL effect requires a belief that the lie-detecting machine is near infallible, we did not find this to be the case.

Nearly all those who disclosed did so during the pre-test stage of the procedure, in the knowledge that their answers would be checked with a 'lie detector'. The reason for

the low number of disclosures in the post-test is unclear, but may have to do with the lack of pressure put on participants in the post-test interview.

Our finding that the BPL effect is present even where there is doubt about the machine's abilities to detect lies leaves open the question of how and why the BPL effect works. Because it appears to be associated only with socially undesirable behaviour (Roese & Jamieson, 1993; Aquinis et al, 1995), it has been suggested that the procedure brings about a motivational shift from self-enhancement or positive self-presentation to one of self-protection, that is, a desire to avoid being seen either to be a liar or being 'out of touch' with oneself, perhaps associated with potential embarrassment, shame, or saving face (Roese & Jamieson, 1993). A 25% error rate, however, would seem to give deceptive individuals a good deal of ambiguity to exploit. It may be that rather than 'self-protection', the presence of the machine gives the individual an 'excuse' to reveal information that they may feel under pressure to keep hidden otherwise due to the processes of self-deception enhancement (von Hippel & Trivers, 2011; Otter & Egan, 2007). Another possibility is that the machine is perceived as providing evidence against the individual, encouraging disclosure in a similar way to confessions in police interview being facilitated by the belief that the police already have evidence of the individual's guilt (Gudjonsson, 2003). If so, findings from the current study suggest that the processes in the bogus pipeline are subtler than the gross indicators examined.

The reality is that the aetiology of the BPL effect remains to be demonstrated. The lack of a relationship with gender, personality factors or suggestibility in our study, however, suggests that whatever mechanisms are involved are likely to be more subtle than are generally assumed. This study was not designed to explore the specific mode of action of the BPL, which should be usefully be examined in future research.



Relevance to Polygraph Testing

A definitive review carried out by the National Research Council (2003) in the United States concluded that the accuracy of polygraph testing is likely to be in the region of 80 – 90%. This is above the 75% accuracy of the bogus lie detector used in this study. Similar to single issue polygraph testing, the focus was on a one important question, which in this case all participants had relevant information. Thus, even if polygraph testing elicits disclosures through the same psychological mechanisms that are at work in the BPL effect (albeit this has still to be shown), the approach does not require psychological manipulation to convince examinees that it is more accurate than it in fact is, contrary to the claims of some critics (Saxe, 1991; Meijer et al, 2008). It also seems simplistic to dismiss the BPL effect as a mere placebo response, as others have argued (e.g., Rosky, 2013).

Limitations

This study's participants were university students who were not facing the consequences of their disclosures in the same way as examinees undergoing polygraphs in criminal justice or employment screening settings. Whether a 75% 'lie detector' would retain its potency in real life testing remains to be determined.

Our power calculation indicated that we needed 155 participants to give us 80% to detect significant effects. We were 14 short of this. Given the strength of the effect sizes observed, however, we do not believe that this shortfall had a meaningful impact.

At the conclusion of the study we did not utilise a post-test assessment to check the extent to which participants accepted that the apparatus was an effective lie detector, or whether claims regarding accuracy were believed. However, whilst participants weren't explicitly questioned about their acceptance of what they were told about the lie detection capability of the apparatus, the significant increase in the proportion of those disclosing in

the BPL groups, suggests that they believed the apparatus was indeed capable of detecting deception. Future research may choose to explore participants pre-existing beliefs regarding the accuracy of lie detector, to examine the potential influence of such preconceptions upon disclosure in all BPL conditions. Another limitation of this study is that there was no group that modelled an actually innocent person, that is one that did not witness cheating, but was accused of it and asked to take a deception detection test. An additional area for future research may be to explore whether the BPL encouraged actually innocent participants or groups of participants to provide confabulated statements.

Conclusion

Given that the BPL effect is seen when the lie detecting machine is only 75% accurate, a level less than the accuracy of techniques such as polygraph testing, it is a misnomer to refer to then BPL as 'bogus'. The effect may depend on a belief that the 'lie detector' has at least some degree of legitimacy, but this does not make it fake – the effect is real and the information it elicits important (Grubin, 2010; Gannon et al, 2014). Rather than downplay its pedigree, perhaps we should now be exploring the more psychologically interesting questions of how the effect works and how it can be enhanced.



References

- Aguinis, H., Pierce, C. A., & Quigley, B. M. (1995). Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: A meta-analytic review. *Basic and Applied Social Psychology*, 16, 515-527.
- Alexander, M. G., & Fisher, T. D. (2003). Truth and Consequences: Using the Bogus Pipeline to Examine Sex Differences in Self-Reported Sexuality. *The Journal of Sex Research*, 40, 27-35.
- Cohen J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd ed. Hillsdale, New Jersey: Erlbaum.
- Cross, T. P., & Saxe, L. (2001). Polygraph testing and sexual abuse: The lure of the magic lasso. *Child Maltreatment*, 6, 195-206.
- English, K., Jones, L., Patrick, D., & Pasini-Hill, D. (2003). Sexual offender containment. *Annals of the New York Academy of Sciences*, 989, 411-427.
- Erdfelder, E., Faul, F. & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Fisher, T. D., & Brunell, A. B. (2014). A bogus pipeline approach to studying gender differences in cheating behaviour. *Personality and Individual Differences*, 62, 91-96.
- Gannon, T. A., Keown, K., & Polaschek, D. L. (2007). Increasing honest responding on cognitive distortions in child molesters: The bogus pipeline revisited. *Sexual Abuse: A Journal of Research and Treatment*, 19, 5-22.
- Gannon, T. A., Wood, J. L., Pina, A., Tyler, N., Barnoux, M. F., & Vasquez, E. A. (2014). An evaluation of mandatory polygraph testing for sexual offenders in the United Kingdom. *Sexual Abuse: A Journal of Research and Treatment*, 26, 178-203.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26-42.
- Grubin D. (2008). The case for polygraph testing of sex offenders. *Legal and Criminological Psychology*, 13, 177-189.
- Grubin, D. (2010) A Trial of Voluntary Polygraphy Testing in 10 English Probation Areas. *Sex Abuse*, 22, 266-278.
- Gudjonsson, G.H. (2003). *The Psychology of Interrogations and Confessions: A Handbook*. Chichester: Wiley.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349-364.
- Kotov, R., Bellman, S.B., Watson, D.B. (2004). Multidimensional Iowa Suggestibility Scale (MISS) Brief Manual. Stonybrook Medicine website. Available at: <http://medicine.stonybrookmedicine.edu/system/files/MISSBriefManual.pdf>. Accessed Jun 10, 2015.
- Meijer, E. H., Verschuere, B., Merckelbach, H., Crombez, G. (2008). Sex offender management using the polygraph: A critical review. *International Journal of Law and Psychiatry*, 31, 423-429



- National Research Council (2003). *The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph*. Washington, DC: The National Academies Press, 2003
- Otter, Z., & Egan, V. (2007). The evolutionary role of self-deceptive enhancement as a protective factor against antisocial cognitions. *Personality and Individual Differences*, 43, 2258-2269.
- Page, R. A., & Moss, M. K. (1975). Attitude similarity and attraction: The effects of the bogus pipeline. *Bulletin of the Psychonomic Society*, 5, 63-65.
- Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: a critical review and meta-analysis. *Psychological Bulletin*, 114, 363-381.
- Rosky, J. W. (2013). The (F)utility of post-conviction polygraph testing. *Sexual Abuse: A Journal of Research and Treatment*, 25, 259-281.
- Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist*, 46, 409-415.
- Sigall, H., & Page, R. A. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 16, 252-258.
- The Telegraph (2012). The awkward truth about lie detectors. <http://blogs.telegraph.co.uk/news/tomchiversscience/100173508/the-awkward-truth-about-lie-detectors/>; last updated 31.7.12.
- Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34, 1-16.
- Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115, 157-168.



The Effect of Questions on Suspects' Perception of Evidence in Investigative Interviews: What Can We Infer from the Basic Literature?

Laure Brimbal¹, Maria Hartwig², and Angela M. Crossman²

¹John Jay College and the Graduate Center, CUNY

²John Jay College of Criminal Justice, CUNY

Abstract

This review presents new ways to conceptualize and investigate how the questions an interviewer asks during a custodial interview can affect a suspect's thought processes. A basic overview of research on perspective taking, common ground, inferences, and psycholinguistics is provided and applied to interview settings. Research on suspects' counter-interrogation strategies and The Strategic Use of Evidence (SUE) technique is summarized serving as a framework within which one can understand the utility of the aforementioned basic research. Issues of interest for practitioners and researchers are discussed based on findings and a conceptual model for factors to take into account in future research is provided.

Keywords: questions, suspects, perspective taking, common ground, counter-interrogation strategies

Over the years, researchers across multiple disciplines (e.g., linguistics, communication research, educational psychology) have examined the various effects of asking questions on people's inferences about these questions (Dillon, 1982). This review is concerned, more narrowly, with the effect of an interviewer's questions on a suspect during an investigative interview. In the interview and interrogation literature, there is a substantial amount of research on the effect of questions. For example, researchers have examined how various forms of leading and suggestive questions affect the reliability of the reports provided by witnesses and victims (Dodd & Bradshaw, 1990; Loftus, 1975; Milne & Bull, 1999). Researchers have also studied the effects of accusatory questions on suspects (Hill, Memon, & McGeorge, 2008; King & Snook, 2009; Narchet, Meissner, & Russano, 2011; Vrij, Mann, & Fisher, 2006). However, as most past research focuses on the construction of the questions (i.e., leading, accusatory), the current review is focused on neutral, non-accusatory questions about various aspects of a crime under investigation. Specifically, it is concerned with how these latter questions affect a suspect's perception about the information held by the

interviewer. This is the core concept of this review. The first portion of the review argues that, in the context of an interview, suspects, particularly guilty suspects, actively seek out information and, critically, draw inferences from the interviewer's questions to guide their counter-interrogation strategies (strategies a suspect uses to appear innocent). The second part reviews different bodies of literature that will help demonstrate how people can draw such inferences and what the likely result of the inferences might be.

This review will utilize and add to the theoretical framework supporting the Strategic Use of Evidence (SUE) technique (Hartwig, Granhag, Strömwall, & Vrij, 2005) because this technique's success hinges on the interviewer's control of the evidence and the suspect's perception of said evidence. However, the conclusions drawn from this review will be applicable to other interviewing techniques that seek to understand or manipulate interviewees' perceptions of evidence. This review will supplement the SUE technique's theoretical framework and help further understand how people process the questions they are asked during a custodial interview. The SUE



technique is a theoretically grounded (Granhag & Hartwig, 2008) and empirically supported (Hartwig, Granhag, & Luke, 2014) interview technique that can be used to help interviewers elicit cues to deception and truth-telling from interviewees (Hartwig, Granhag, Strömwall, & Kronkvist, 2006) and can also possibly be used to elicit information from suspects (Luke, Dawson, Hartwig, & Granhag, 2014a; Luke, Shamash, Hartwig, & Granhag, 2014b). To inform the argument that suspects use counter-interrogation strategies, we review relevant interviewing literature, beginning with research on suspects' counter-interrogation strategies, as it is the basis for the theoretical framework for the most relevant interviewing techniques. We explain the link between suspects' perception of the evidence and the counter-interrogation strategies they employ. We then examine how the SUE technique is related to suspects' counter-interrogation strategies, and how the technique can be used to examine how interviewees critically

evaluate questions to guide their counter-interrogation strategies.

To develop an adequate framework of how people draw inferences from the questions that they are asked, we draw from different but equally important fields (see Figure 1). From a broad view, perspective-taking research is relevant, as it explains how people draw inferences about others' knowledge. Similarly, common ground research tackles the question of how people use knowledge that is familiar to both them and their interlocutor when communicating. From a narrower perspective, psycholinguistic research on how people encode and process questions is important as well, as it gives us an understanding of what information is retrieved from the wording of questions before more complex and strategic thought processes are involved. The review concludes with a summary of the literature, the proposed model, and related predictions and suggestions for future research.

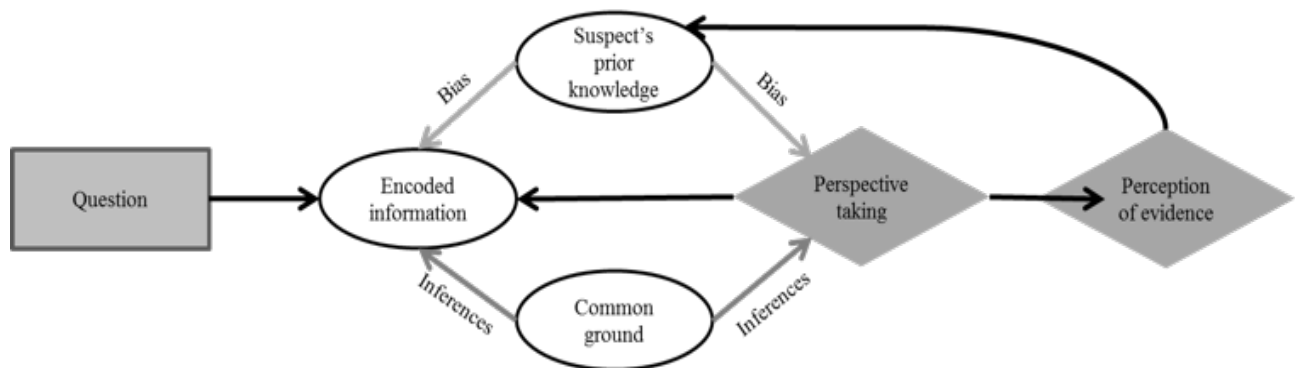


Figure 1. The framework detailing the relationship between information and processing described in this review: As they are asked by the interviewer, questions are processed by the suspect. The information from these questions that is encoded can be biased by the suspect's prior knowledge and affected by inferences from the common ground knowledge between them and the interviewer. If the suspect attempts to take the interviewer's perspective, they will base their judgment of what information the interviewer has on the encoded information from the questions they were asked. Their perspective taking activities can also be affected by inferences about common ground knowledge and biased by their own prior knowledge. Finally, the suspect perspective taking activities result in their perception of the evidence held by the interviewer.



Research on Interviewing

Suspects' Counter-Interrogation Strategies

Research on suspects' counter-interrogation strategies illustrates that innocent and guilty suspects approach an interview in different states of mind (Granhag & Hartwig, 2008). Typically, innocent suspects report having no strategy to convince the interrogator of their sincerity. If they do report a strategy, it tends to involve verbal forthcomingness. For example, in one study, the principal strategy reported by innocent suspects was to simply tell the truth as it happened (Strömwall, Hartwig, & Granhag, 2006; see also Hartwig, Granhag, Strömwall, & Doering, 2010). It seems that innocent people believe that if they simply provide a full account of what happened, their innocence will be apparent and the interrogator will 'see' that they are telling the truth. This belief in the visibility of innocence is supported by research on the phenomenology of innocence and different factors that have been empirically shown to make innocent suspects falsely confess to crimes (Kassin, 2005).

More broadly, people have attributed innocents' tendency to assume they are transparent to their belief in a just world (Lerner, 1980; Lerner & Simmons, 1966) and the illusion of transparency (Gilovich, Savitsky, & Medvec, 1998). Lerner's phrase 'belief in a just world' refers to people's need to believe that good things happen to good people and that the world is a just place. Although people vary individually in the need for this belief, it is thought to be a general human tendency (Hafer & Bègue, 2005; Lerner & Miller, 1978). Belief in a just world could be causing innocent suspects to be overly confident that because they are good, innocent people, they will be treated well – and be believed. The illusion of transparency is the tendency to believe that our internal states are more apparent to others than they actually are (Gilovich et al., 1998). This would lead innocent suspects to believe, perhaps erroneously, that their innocence will be clear to the interviewer. The illusion of transparency extends well beyond the realm of deception (Epley, Keysar, Van Boven, & Gilovich, 2004; Gilovich, Medvec, & Savitsky, 2000) and will be discussed in more depth in the section on perspective taking below.

In contrast, guilty suspects engage in complex decision-making processes. In general, guilty suspects are motivated to conceal the truth about their activities to try to convince the interviewer that they are innocent. Guilty suspects must provide a deceptive statement that minimizes the risk of being caught in a lie, in case the interviewer has information about the crime. For example, a thief being interviewed about a break-in he just committed may be asked what he was doing during that time. He might answer that he was walking around the neighborhood where the break-in occurred. This answer is vague in that it omits the fact that he was in the building, but does not eliminate the option to be forthcoming if probed further. Indeed, research shows that guilty suspects are mainly concerned with this type of information management during interviews (Colwell, Hiscock-Anisman, Memon, Woods, & Michlik, 2006; Hines et al., 2010).

Information Management Theory (IMT; McCornack, 1992) suggests four deceptive ways a person can manipulate information. People can: falsify information, avoid mentioning all information, be ambiguous in their statements, and present irrelevant information. In interviewing settings, liars report using strategies such as keeping their story simple (thus withholding details that could incriminate them), avoiding or denying incriminating details, and using avoidant counter-interrogation strategies (Hartwig, Granhag, & Strömwall, 2007; Strömwall et al., 2006). These results have been supported by survey data asking incarcerated criminals what strategies they used when lying (Granhag, Clemens, & Strömwall, 2009). These differences (e.g., between guilty and innocent suspects' reasoning, planning and strategies) form the theoretical background for the SUE technique (Hartwig et al., 2005; 2006).

The Strategic Use of Evidence (SUE) Technique

The SUE technique is an interview technique based on instrumental mind reading (Perner & Kuhberger, 2005). That is, interviewers are trained to imagine themselves in the suspect's situation to better predict their thought processes (Granhag & Hartwig, 2008; see Figure 2). Thus, for example, an interviewer using the SUE technique is concerned



with a suspect's perception of the evidence. It is hypothesized that guilty suspects use the amount of information they believe their interviewer has to guide their interview strategy should be to minimize the chance of getting caught (Granhag & Hartwig, 2015).

The SUE technique consists of several phases, two of which are relevant to this review (for an in-depth discussion of the technique, see Granhag & Hartwig, 2015). The interview begins with a series of free recall prompts and open-ended questions. These al-

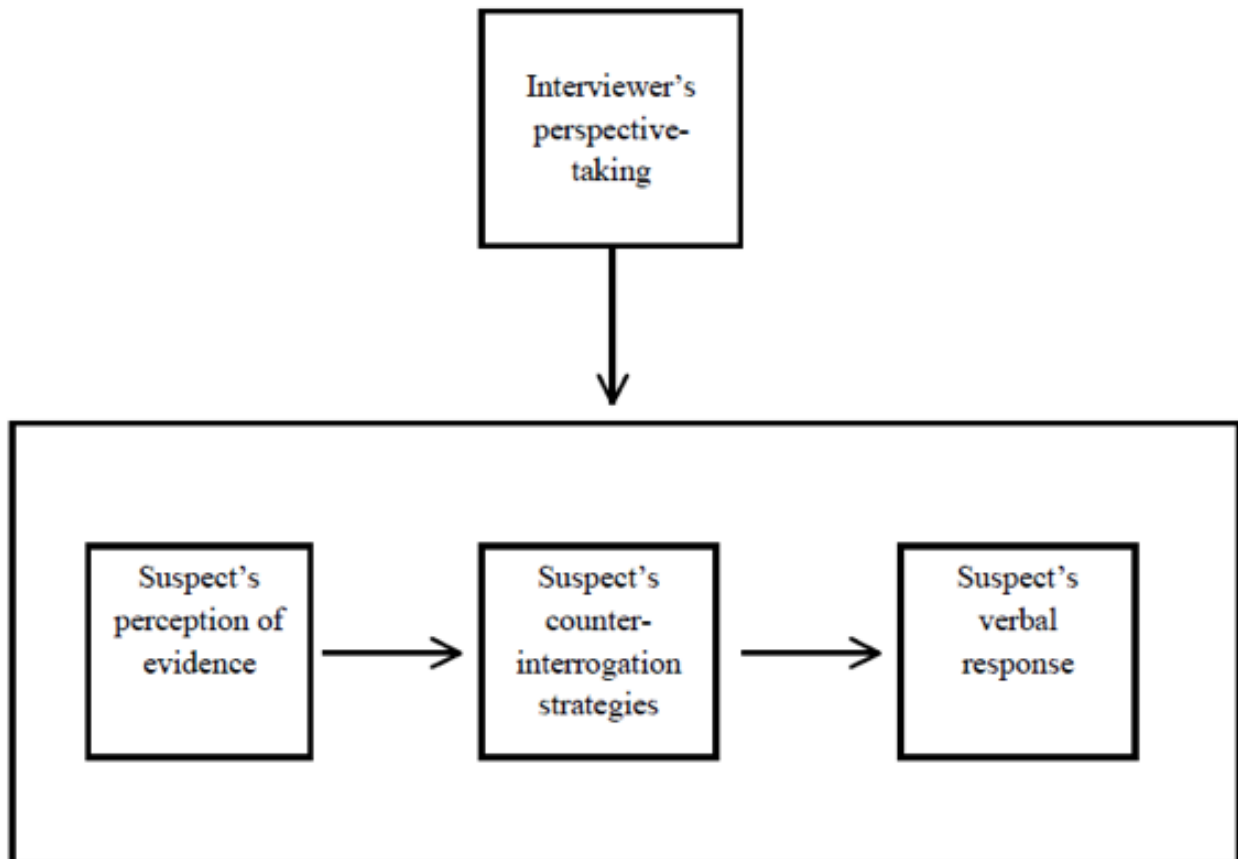


Figure 2. The relation between suspect-related principles and perspective-taking on behalf of the interviewer. Reprinted from *The Strategic Use of Evidence (SUE) technique: A conceptual overview* (p. 14), by P. A. Granhag & M. Hartwig, 2015, John Wiley & Sons. Reprinted with permission.



low suspects to give an account of activities within the time frame of the crime they are under suspicion of having committed (Phase 1). This phase should result in omissions from guilty suspects (Granhag, Strömwall, Willén, & Hartwig, 2013). As stated above, guilty suspects typically chose avoidant counter-interrogation strategies, such as omitting incriminating information, as this is less risky than outright contradicting the facts. Innocent suspects, on the other hand, should be forthcoming regardless of the phase of the SUE technique. Phase 2 involves the interviewer asking questions about specific incriminating evidence (without disclosing it). The purpose of this phase is to induce a suspect to commit to a specific account of their activities (Hartwig et al., 2005). For example, if the police have video footage of a suspect entering the building where a break-in occurred, the interviewer will use phase 2 to ask the suspect if they have ever been in the building in question, or if they know anyone who lives in the building, etc. In response to these types of questions, guilty suspects typically give statements that contradict the evidence (see Hartwig et al., 2014). Once a suspect has committed to an account, the interviewer presents the suspect with the evidence they have, thus moving on to phases 3 and 4, the disclosure phases¹. Accordingly, the first two phases of the SUE technique entail having the interviewer place themselves in the shoes of the suspect. Research on this technique has suggested that a suspect's perception of the evidence guides their counter-interrogation strategies (Luke et al., 2014a).

The Effect of Suspects' Perceptions of the Evidence

As explained above, guilty suspects typically use avoidant counter-interrogation strategies and avoid mentioning incriminating information (Granhag & Hartwig, 2008). Certain circumstances however can prompt guilty suspects to be more forthcoming with information; thus resembling innocent suspects. Recently, Luke et al. (2014a) hypothesized that guilty suspects' verbal strategies are influenced by their perception of the amount

of evidence against them (henceforth, perception of the evidence). The default strategy for a guilty suspect seems to rely on avoiding or denying incriminating information. However, if they anticipate that there might be evidence against them, they may switch to a more forthcoming verbal strategy, to avoid the risk of providing a statement that conflicts with the facts held by the interviewer. Their prediction was supported: Guilty suspects who were not told anything about the possible existence of evidence against them were prone to withhold information. In contrast, guilty suspects who were informed of the possibility of evidence against them (without being told exactly what this evidence was) chose to employ a highly forthcoming strategy, volunteering large amounts of information (Luke et al., 2014a).

Thus, it seems that a guilty suspect's perception of the evidence is an important dimension when determining which strategy employed during an interview, which in-turn affects the amount of information they withhold or deny (Granhag & Hartwig, 2015).

Since the SUE technique exploits a liar's strategy to withhold or deny incriminating information, understanding that strategy can be of considerable importance (Hartwig et al., 2014). By generating implausible or incriminating denials, liars' verbal strategies may distinguish them from truth-tellers. Furthermore, interviewers in some settings might want a suspect to be forthcoming, regardless of guilt. For example, in intelligence gathering contexts, a primary challenge is to elicit information (Evans, Meissner, Brandon, Russano, & Kleinman, 2010). Hence, an interviewer who understands a liar's choice of strategy is in a better position to accomplish their goal, whether that goal is detection of deception or information elicitation (Luke et al., 2014a).

A prerequisite for the SUE technique to be employed effectively is that suspects are unaware, or at least uncertain, of the evidence collected in an ongoing investigation. Uncertainty in interpersonal interaction has been shown to motivate people to seek out in-

¹ The disclosure phases are not directly relevant to this review and will not be discussed further.



formation to resolve it (Afifi & Weiner, 2004; Afifi & Weiner, 2006). We hypothesize that the questions asked using the SUE technique are a main source of potential information acquisition for suspects, due to their ignorance of evidence. This is instrumental for the technique's success. Thus, suspects should be motivated to derive information about what the interviewer knows from the questions that they are asked. If the questions imply that the interviewer holds evidence against a suspect, the interviewer might, without realizing it, influence their counter-interrogation strategies. Hence, it is critical to study the mechanisms that underlie these information-seeking activities. Broadly speaking, suspects might attempt to take the interviewer's perspective by processing the questions and drawing inferences from them to inform their counter-interrogation strategy. For this reason, we now examine previous psychological research on perspective taking, questions and inferences.

Previous Work on the Effect of Questions on Perception of the Evidence

To investigate the effect of questioning about specific themes, Brimbal, Hartwig, and Luke (2013) conducted a study in which participants were asked to imagine themselves as members of a terrorist organization being interviewed by the police. They were randomly assigned to read different versions of the interview transcript. It consisted of an open-ended question (asking the suspect to walk them through their day) to which the suspect answered that they had been at the train station only, the post office only, neither location, or both. The interviewer then asked one last question probing for more detail about the post office, which the suspect did not answer. Participants were then asked to report how much knowledge they thought the interviewer had against them (the suspect). The results showed that merely asking about a topic implied knowledge about that topic. Participants also thought that the interviewer knew less about a topic that the suspect had previously mentioned compared to a topic that the suspect had not mentioned (Brimbal et al., 2013). These findings illustrate that perceptions of the interviewer's knowledge vary dynamically with the questions the interviewees are asked.

Research on Perspective Taking

Perspective taking is the cognitive capacity to consider the world from other viewpoints and 'allows an individual to anticipate the behaviour and reactions of others' (Davis, 1983, p. 115). It may help to provide an explanation for how people actively perceive others' thought processes (Epley & Waytz, 2010). This can be distinguished from impression formation, which deals with the characteristics people ascribe to others based on their external features and behaviours (Ambady & Rosenthal, 1992; Bierhoff, 1989; DeCoster & Claypool, 2004; Fiske & Neuberg, 1990; Fiske & Taylor, 2013; Greenwald & Banaji, 1995). Perspective taking can also be distinguished from inferences about others' internal states, in the context of empathy and prosocial behaviour, such as helping (Coke, Batson, & McDavis, 1978), which places an emphasis on how people view others' emotions and feelings. We are mainly interested in how people draw inferences about others' thoughts and knowledge, as this is most relevant to suspects' counter-interrogation strategies.

Perspective taking is important for an interviewer, as discussed above, particularly when using techniques such as the SUE, and others not discussed here (e.g., Granhag, Cancino Montecinos, & Oleszkiewicz, 2013; May, Granhag, & Oleszkiewicz, 2014; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014), that hinge on anticipating what an interviewee is thinking. However, the interviewee's ability and motivation to take the perspective of the interviewer during questioning to draw inferences about what they know is also an important factor in their exchange. With this in mind, we discuss the underlying mechanisms of perspective taking. We review different theories of how people evaluate others' knowledge and explore how this can be used to explain how interviewees infer knowledge from an interviewer's questions.

Perspective taking ability develops from Theory of Mind (Astington, Harris, & Olson, 1988; Lewis & Mitchell, 1994; Mitchell &



Riggs, 2000; Premack & Woodruff, 1978; Wellman, Cross, & Watson, 2001), which explains how a person can know and understand that others have beliefs, independent from their own (Flavell, 1986; Perner, 1991). Much of the research on Theory of Mind concentrates on when and how children develop Theory of Mind, as it is not a faculty that is innate, but one that is developed through socialization (Flavell, 2000; Perner, Ruffman, & Leekam, 1994; Piaget, 1965). Theory of Mind research is not limited to children and can be applied to adult social cognition as well (Apperly, Samson, & Humphreys, 2009; Keysar, Lin, & Barr, 2003). For adults, however, beyond Theory of Mind, researchers also investigate the likely related skill of perspective taking.

Several theories have been developed to explain the mechanisms behind how people can predict others' thoughts. Floyd Allport (1924) was the first to formally investigate the phenomenon, though he coined it social projection. He surveyed college students about their own cheating behaviour and their beliefs about others' cheating behaviour. He concluded that others are assumed to be very similar to ourselves (Katz & Allport, 1931). Since that time, general theories about perspective taking can be encompassed under two large umbrellas: simulation and theory driven. A simulation driven approach was initially devised by Abel (1948), who posited that we predict what others think by projecting ourselves into others' situations and asking ourselves what we would do or think in their situation (Maccoby, 1980). In opposition to this view, Chandler (1976) suggested that trying to figure out what we ourselves would do is not actually perspective taking and that real perspective taking requires ignoring our own point of view. Higgins (1981) bridged the two perspectives by differentiating situational and individual role taking. Situational role-taking is similar to what Abel suggested, while individual role taking is closer to Chandler's conception. This type of simulation is supported by research that shows that people generally construct estimates of others' knowledge or emotional states by imputing their own knowl-

edge or likely emotions to those others (Abel, 1948; Epley & Waytz, 2010; Karniol, 1990; Keysar, 1994; Nickerson, 1999; Sarbin, 1954; Van Boven & Loewenstein, 2003). Though individual role taking appears to be the ideal for perspective taking, it has been hypothesized that situational role taking is the best reference that people have for predicting others' thoughts because the only mind they have access to is their own (Wegner, 2005).

The theory driven perspective ignores the role of self-knowledge and posits that individuals make predictions about others based on three types of information: information about the situation, general information about how people react in similar situations, and specific information about the person whose thoughts and emotions they are trying to predict (Gopnik, 1993; Karniol, 2003; Malle, 2005; Nisbett & Wilson, 1977; Saxe, 2008). For instance, Karniol (1990) introduced the idea that we use transformation rules to make decisions about what others are thinking and feeling. In her model, Karniol posits that people then use the transformative rules to fill in with declarative knowledge (e.g., facts, information, and ideas) about such situations (Schank, 1982; Schank & Abelson, 1977) and people's individual differences, which often depend on our understanding of others' goals (Foss & Bower, 1986; Locke, 2000). This illustration of theory driven perspective taking implies that we use rules and knowledge to systematically evaluate others' perspectives (Gopnik & Wellman, 1992; Karniol 2003). This perspective further suggests that we base our inferences on information that we gather through our knowledge of the world, stereotypes, observations of the specific individual's behaviours, or knowledge about someone specific, indirectly acquired (Ames, 2004a, 2004b; Gopnik & Wellman, 1994; Karniol, 1990).

To resolve the differences between theoretical frameworks in the field, Epley and Waytz (2010) proposed a dual process model. According to this model, when trying to mind read, one typically starts off with one's own perception of the world (Dawes & Mulford, 1996; Krueger & Acevedo, 2005). This is illus-



trated by the fact that our inferences about others' minds (i.e., goals, intentions, attitudes, knowledge, etc.) are habitually biased towards our own (see Alicke, Dunning, & Krueger, 2005). After this egocentric simulation, one uses theoretical analogies in a controlled manner to correct for these potential biases.

Empirical evidence supports this dual process in developmental research. Children begin their lives without Theory of Mind, and consider everything from an egocentric point of view. As they grow, they develop more 'mature' reasoning, stepping outside of the constraints of their own biased frames of reference to contemplate things from another's point of view (Moore, 2005). Birch and Bloom (2007) found that adults continue to show signs of egocentric anchoring when completing a task that is similar, but more sophisticated than that used to assess Theory of Mind. Furthermore, neuroimaging research showed that, when asked to imagine that they were someone else (perspective taking), participants blurred the distinction between the self and another (Ames, Jenkins, Banaji, & Mitchell, 2008). Also, research indicates that adults overestimate how close others' attitudes and feelings are to their own (Krueger & Clement, 1994; Ross, Greene, & House, 1977), and think their internal states are visible to others (Gilovich et al., 1998). Despite showing egocentric tendencies, however, it has been shown that adults are better able to control these tendencies than children (Epley, Morewedge, & Keysar, 2004).

According to the dual process model, egocentric tendencies are automatic, while suppressing them is a controlled, effortful process, presumably explaining adults' greater facility than children in controlling them. Epley and Waytz (2010) propose that a more complex form of perspective taking is using information about others to adjust our initial egocentric interpretations of what they say (Epley et al., 2004; Keysar & Barr, 2002). In support of this model, Converse, Lin, Keysar, and Epley (2008) found that positive mood affected perspective taking ability. This is in line with research which suggests that happiness lowers people's ability to use executive control functions (Oaksford, Morris, Grainger, & Williams, 1996; Phillips, Bull, Adams, & Fraser, 2002) and the tendency to rely on elaborate

processing of information (e.g., Bless & Igou, 2005; Forgas, 1995; Isen, 1984). Indeed, making people happy made them more likely to have egocentric views of what others believed instead of using the more cognitively effortful strategy of applying their knowledge of others to make these judgments.

Apperly, Riggs, Simpson, Samson, and Chiavarino (2006) investigated whether encoding other people's beliefs was automatic in two ways. They manipulated whether participants were explicitly told to monitor other's beliefs and measured participants' reaction times to probe sentences about said beliefs. To do this, participants were shown a video depicting two characters, two boxes and an object. The object was placed in one of the boxes in a room and participants were told to keep track of which box it was in. A female character looked into the boxes and indicated where the object was hidden and then left the room. A male character then came in and switched which box the object was in, changing reality (i.e., where the object actually was) and the female character's belief from true to false (her initial belief became false when the object was moved). Since the change in the female character's belief was irrelevant to the participants' task, the participants should only track the female's belief if they do so automatically. Participants were presented with either a reality probe ('It's true that it's in the box on the right') or a belief probe ('She thinks that it's in the box on the left'; Apperly et al., 2006, p. 842) and reaction times for their responses (either yes or no – indicating whether the probe was true) to these probes were measured. Reaction times were used to assess whether participants were faster or slower to respond to the belief probe, implying respectively that they had either already processed the female character's belief or not. The authors found that, unless expressly instructed to follow the female character's beliefs, participants responded more slowly to the belief probes than to the reality probes, indicating that they were not processing her beliefs automatically.

Several other studies have provided evidence for the automaticity of egocentric perspective taking. For example, when people are asked to assess others' intentions, while having to conduct a second task (Lin, Keysar, & Epley, 2010) or are asked to respond quick-



ly (Epley et al., 2004), they tend to use their knowledge about others less. Furthermore, stress has been shown to change the way people make predictions about others' thoughts (Karniol, 1990). In the same vein, research on negotiations has found that people, even when trained in negotiations, do not spontaneously take others' perspectives, but may learn to do so (Neale & Bazerman, 1983; Thompson & Hastie, 1990). Indeed, MBA students learning how to negotiate were more successful when they were told to take the perspective of the other negotiator rather than when they were given no instruction. The MBA students reached more deals with more individuals, had more collective gains and used more creative solutions when they were given perspective taking instructions. Surprisingly, perspective taking was a more effective means of negotiation than empathizing with a partner (Galinsky, Maddux, Gilin, & White, 2008).

Epley and Waytz's dual process model (2010) is interesting from an applied perspective because being interviewed can be a cognitively demanding task. It is subsequently important to consider factors (here cognitive load, stress, mood, etc.) that can affect the way suspects take the perspective of the interviewer. This model also illustrates the different conditions that might affect suspects' interpretation of the questions they are asked and the different types of biases they might be subject to. As discussed above, it can be cognitively demanding and particularly precarious to take someone else's perspective and uncover what information they have, because others' perspectives can sometimes be entangled with one's self-perception. Epley and Caruso (2008) reviewed three common reasons why people fail to take someone else's perspective. The first is that, because it is hard to move away from our own egocentric perspective, people often do not even attempt it. Furthermore, one's initial egocentric perspective can bias any future instrumental mind reading, resulting in a failure to adjust accurately to someone else's perspective. Finally, it is possible that even if one manages to detach from one's own egocentric biases, knowledge about others is wrong and thus one fails to accurately adjust to their perspective.

When we are trying to predict what someone else is thinking or feeling, we refer

to our own schemas of how we would deal with the situation that s/he is in (Abel, 1948; Karniol, 1990; Sarbin, 1954). This process of putting oneself in another's mind, yet using one's own knowledge to draw inferences about the other's thoughts could lead to confusion over where that knowledge came from. These issues are more commonly known as source monitoring errors, which is the failure to attribute cognitions to their proper source of origin (Johnson, 1988; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981; Mitchell & Johnson, 2000, 2009; Yonelinas, 2002). Source monitoring has been a concern for researchers in the past, as it has been shown to affect the accuracy of eyewitness reporting (Dobson & Markham, 1993), persuasion (Henkel & Coffman, 2004; Underwood & Pezdek, 1998), aging (Hashtroudi, Johnson, & Chrosniak, 1989) and amnesia (Mitchell & Johnson, 2009). When considering failures to take an interviewer's perspective, the content of an interviewee's answers could be misattributed to an interviewer's questions. This would lead interviewees to use their own responses as a basis for their inferences about the interviewer's knowledge. For example, if an interviewee is asked where they were in a building at a specific time, they might respond that they were in the mailroom. If later questioning involves this mailroom and the interviewee does not remember that they mentioned the mailroom before being questioned about it, they might infer that the interviewer has more knowledge about their activities than they in fact have.

Research on Common Ground

Common ground is defined as a type of shared information: 'The common ground between Ann and Bob, for example, is the sum of their mutual knowledge, mutual beliefs, and mutual suppositions' (Clark, 1992, p. 3). The common ground is contrasted with the privileged ground (Clark & Carlson, 1981; Clark & Marshall, 1981) that refers to what only one interlocutor knows. Common ground is an intricate component of language comprehension (Clark & Carlson, 1981; Clark & Marshall, 1981). For example, when a speaker uses a term that refers to a very common object such as 'the stop sign', this could be considered ambiguous. Nonetheless, knowing what common knowledge there is between the speaker



and the listener will help the listener resolve which stop sign the speaker is referring to. Indeed, it is difficult for people to communicate effectively if their partner is not exposed to the same information as they are (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995).

Research on common ground extends perspective taking research, as it reflects mutual knowledge that might be shared in an interview room. For example, assuming a suspect is guilty, an interviewer might believe that mutual knowledge exists (e.g., facts of the crime), an assumption that is false when the person is innocent. Guilty suspects in contrast, must determine what the common ground is (i.e., what the interviewer knows, that they also know because they committed the crime). Common ground is especially important in such a domain because both parties might be attempting to conceal information in the common ground or strategically placing information in the common ground to track what information is shared. Research on common ground does not typically encompass this type of strategic communication situation, but rather focuses on regular conversations. However, it is important to understand how we use common ground knowledge in normal discourse processing to infer how people might reason about it in a suspect interview.

The typical paradigm used to investigate common ground evolved from a 1966 study by Krauss and Weinheimer where participants were instructed to decide which of six cards a confederate had in front of him or her. The confederate communicated instructions via telephone to restrict the common ground knowledge between him/her and the participants. More recently, researchers used displays of objects, some of which were visible to a confederate and some that were not. Some of the hidden objects (those in the participants' privileged ground) had matching descriptions (referred to as competitors) with those in the common ground (e.g., one golden star is visible to both participant and confederate and one gold star –the competitor– is visible only to the participant). The confederate provided instructions on how to move the objects that were in the common ground. Using eye-tracking software, researchers tracked whether participants acknowledged the competitor in their privileged ground and/or how long it took

them to discount these objects (Cooper, 1974; Keysar, Barr, Balin, & Brauner, 2000; Tanenhaus, Spivey, Eberhard, & Sedivy, 1995). This research shows that people at times use an egocentric heuristic; meaning they disregard common ground knowledge and consider only their own. Findings also show that since this heuristic can lead to errors in interpretation of language, people can also discount common ground when they are interpreting engaged in a dialogue.

Several theoretical models were developed to account for when and how people use common ground knowledge. Horton and Keysar (1996) proposed a Monitoring and Adjustment model for language production (see also Dell & Brown, 1991) that is similar for language comprehension (Keysar, Barr, Balin, & Paek, 1998). They hypothesized that people only use common ground knowledge if they make a mistake following an unrestricted memory search. This can be compared to a restricted search in which common ground knowledge restricts the memory structures that a person searches when they intend to produce an utterance. More specifically, this model posits that we first rapidly conduct an unrestricted search that has no concern for mutual knowledge, but we then complement it by a system that monitors and adjusts according to what is in the common ground. The two processes take place simultaneously (McClelland, 1979), but the monitoring does not go as fast as the search and thus is sometimes unable to stop mistakes from happening (i.e., someone says or understands an utterance as something that is not within the common ground).

On the other hand, the constraint-based model of common ground utilization (Hanna, Tanenhaus, & Trueswell, 2003) postulates that we do indeed use common knowledge initially, which restricts our interpretation of information. Hanna et al. found support for their model by tracking participants' eye movements while they performed the task described above. In one condition, one of the objects matched the target object but was only in the participant's privileged ground (i.e., visible to only the participant). Researchers found that very quickly after the description of the target object was given, the participants discontinued looking at the similar object that



was in the privileged ground. These results were replicated (Nadig & Sedivy, 2002) in more realistic, externally valid environments, such as helping a cook locate ingredients (Hanna & Tanenhaus, 2004). Researchers found that common ground did not account for all of the processing restriction; however there was no evidence to support an initial phase of processing that disregarded completely the common ground (Keysar et al., 2000). Hanna and her colleagues' constraint based model (2003) takes the form of a probabilistic model, where different constraints come from different sources of information and are weighed according to their respective saliences and reliability. This means that all possible interpretations of a sentence have the same probability of being selected as true when encoding begins and they are individually discounted based on the constraints (Hanna et al., 2003). Common ground is one of many factors that can affect the discounting of options and interpretation of a sentence.

Lastly, with his anticipation integration model, Barr (2008) criticized the research described above as flawed because the experimental paradigms used have ambiguities that cannot be resolved through linguistic evidence. Indeed, in the typical common ground experiment, the instructions are not given in a manner that is comparable to natural dialogue. For example, in a regular conversation, a person can typically resolve the ambiguity of whether the confederate is referring to the gold star in the common ground or the competitor. In his studies, Barr (2008) attempted to account for the possibility that people anticipate that others will be referring to things in the common ground. He distinguished anticipation without integration from anticipation with integration. The former was associated with autonomous activation as the listener would anticipate the common ground, but not utilize it when processing discourse. The latter was associated with constraint-based models that posit use of the information as it is learned to constrain their interpretation of language. To test this model, Barr (2008) conducted three studies that found support for anticipation without integration, as people were not integrating their knowledge of the common ground into their responses.

Overall, the literature suggests that people can integrate common ground knowledge into their communication in order to resolve ambiguities, but this is cognitively costly and entails constant perspective taking (Lin, Keysar, & Epley, 2010; Rossnagel, 2000). As people are not always motivated or cognitively able to take another's perspective and thus monitor common ground, they often will rely on privileged information when speaking (Barr & Keysar, 2006; Brown & Dell, 1987; Ferreira & Dell, 2000). This means that in a complex interaction such as an interview, where common ground is purposefully convoluted, it might be especially difficult for a suspect to track and utilize common ground when answering the interviewer's questions.

Research on Inferences

Pragmatic Inferences

The effects of asking questions are typically studied in the context of interpersonal interactions that are ruled by Gricean principles (Grice, 1975). Grice's four maxims of cooperative communication are: quantity, quality, relation, and manner (Grice, 1975). These indicate respectively that the speaker should speak as much as necessary to be understood (but not more), s/he should be truthful, speak about things relevant to the situation and finally that the contribution be easy to understand. These principles are claimed to be necessary to cooperative and efficient conversations and guide research in linguistics. In interviewing settings, when both interviewer and interviewee might be purposefully deceptive, these principles do not always apply (McCornack, 1992). On the one hand, linguistics research supports the idea that asking questions pragmatically implies that one does not know its answer (Belnap, 1966; Dillon, 1982) and the person they are asking does and should answer (Goffman, 1974; Schegloff & Sacks, 1973). On the other hand, asking questions in certain situations (e.g., interrogations) can suggest that the questioner could have a reason for asking the question (Wegner, Wenzlaff, Kerker, & Beattie, 1981) such as having more information than they would like to acknowledge.



Pragmatic inferences are common sense inferences that people draw from what is said, including questions. Typically, they are studied at a structural level where the construction or content of a sentence pragmatically implies a fact. For example, saying that person A and person B are friends pragmatically implies that person A and person B know each other. In terms of questions, the presupposition is the information that is pragmatically implied to be true in order for that question to be asked (Belnap, 1966). For example, if a suspect is asked, 'why were you in the store?' this pragmatically implies that they were indeed in the store. This review is primarily concerned with the pragmatic implications questions asked and not the construction of the questions (unlike the research on leading questions). Yet it is important to consider the presuppositions of questions as they likely guide suspects about the 'theme' of questioning. The presuppositions can inform a suspect about what, specifically, the interviewer is asking about and thus what information they may or may not know. Furthermore, research shows that ambiguous questions can lead a questioned person to rely on external factors to understand them (Strack, Schwarz, & Wänke, 1991). This will force the individual to evaluate contextual cues to explain why they are being asked the question. In these ways, pragmatic inferences are an important component of the dynamics in an interrogation situation.

Elaborative Inferences

Elaborative inferences are inferences drawn about the macrostructure of a discourse (Singer, 1994). The macrostructure of a discourse regards its main idea or the overarching theme that a speaker is addressing and can be conceptualized as the top of a hierarchical network of micropropositions (Kintsch & van Dijk, 1978). Micropropositions are the simplest propositions that inferences (such as pragmatic inferences) can be drawn from. This is relevant to the current review as each question an interviewer asks can be considered a microproposition, while each theme of questioning (i.e., each piece of evidence that the interviewer is interested in and develops a line of questioning about) can be conceptualized as a macrostructure. The inference theme hypothesis (Singer, 1994) posits that people priori-

tize drawing inferences about thematic ideas rather than ideas that are peripheral (Cirilo, 1981; Walker & Meyer, 1980). This means that people will pay more attention to consistent themes of questioning and not elements that seem less central to the questioning.

Inferences also play a role in the concept of common ground (Eysenck & Keane, 2010; Stalnaker, 1978, 2002; Clark & Marshall, 1981; Clark & Schaeffer, 1989; Kass & Finin, 1988). Speakers typically incorporate their beliefs about other's knowledge into their language production processes and thus often use the common ground to make pragmatic inferences (Horton & Gerrig, 2005). For example, if an interviewer says that they will be asking questions about a crime committed in the interviewee's apartment building, the fact that a crime was committed becomes common ground and when the interviewer refers to the 'crime', the interviewee can pragmatically infer that the interviewer is referring to the neighbour's murder. This is important because the pragmatic inference that a suspect makes about questions the interviewer asks could be about information that is (e.g., the reason that the suspect is being interviewed) or is not common ground (e.g., evidence that the interviewer has against the suspect but has not disclosed yet). When the interviewer specifically presents information or evidence to the interviewee, there should be less ambiguity about what the interviewer knows. Thus we predict, generally, that when information is in the common ground (that is, when both suspect and interviewer have discussed it previously), then guilty suspects will expect Gricean principles to apply. They will expect that the interviewer is asking questions because they have less evidence and need the guilty suspect's statement to build a case against him. However, when information is not in the common ground, then guilty suspects should infer more knowledge on the part of the interviewer.

Research on elaborative inferences considers both factors that affect how people draw inferences and what inferences are primarily drawn about. Long, Golding, and Graesser (1992) found that people drew elaborative inferences to determine people's superordinate goals (i.e., the motivation and goal behind someone's actions). Translated into in-



interviewing settings, this means that suspects might draw inferences about the interviewer's goals, regardless of guilt. On the other hand, people appraise the implications of events that have serious consequences (e.g., dying in a plane crash) more accurately than those with trivial consequences (Goetz, 1979). Thus, consequences of the interview should impact the inferences suspects draw: If the interviewee thinks that the ramifications of the outcome of the interview could be important, they will be motivated to be accurate with their inferences. In addition, processing capacity seems to be related to the manner and accuracy with which people draw inferences. Verbal working memory capacity is related to how accurate people are when drawing elaborative inferences (Dixon, Lefevre, & Twilley, 1988; Masson & Miller, 1983). Whitney, Ritchie, and Clark (1991) found that verbal working memory capacity was a factor in how people made inferences and Calvo (2005) found that verbal working memory capacity helped participants integrate meaning from inferences. Thus, as with perspective taking, it seems that cognitive capacity plays a part in the interpretation of the subtleties of language and is likely relevant to an interviewee's ability to infer information from the questions of an interviewer.

Research on Questions

How are Questions Processed and Encoded?

Research on common ground and psycholinguistics rarely accounts for the cognitive processes of language production and comprehension (Pickering & Garrod, 2004). As the current review is concerned with a very specific type of dialogue (i.e., an interview composed almost entirely of questions and their ensuing answers), literature on question processing is reviewed to provide a clear model of information processing. In particular, the literature on questions is reviewed, as these are the source of information of interest for suspects. The focus is on how people cognitively process questions and how information is typically encoded when being asked questions in an attempt to understand what people are likely to be focusing their attention on and what information they are obtaining from questions.

Questions can be viewed as composed of two components: the presupposed information and the focal information (Graesser, McMahon, & Johnson, 1994). The presupposition is defined as the relationship between two clauses (e.g., A and B) where clause A needs to be true for clause B to be considered either true or false (Strawson, 1952). By analogy, the presupposition of a question is the information in this question that is considered given and is considered common ground (Stalnaker, 2002) or the knowledge that is shared by both parties (i.e., the questioner and the questioned) during a dialogue. A simple example of this is if an interviewer asks: 'Where did you go when you were in the building?' The presupposition of this question is that the suspect was in a building that both the suspect and the interviewer know of. On the other hand, the focal point of a question is the part of the question that requests information. Research on flaws in sentence comprehension shows that people do not typically analyze the information from presuppositions critically. People do however, evaluate carefully the information that is the focus of the question (Reder & Cleeremans, 1990) and most attention is paid to the information that the questioner wants to know (Erickson & Mattson, 1981). In the same question as above (i.e., 'Where did you go when you were in the building?'), the focus of the question is the portion asking where the interviewee went and less attention should be spent on the specification of the location asked about. Expectation about what is going to be asked about is used as a source to inform the interpretation of the information not in the question's focus (Park & Reder, 2004). As our cognitive capacity limits our linguistic processing capabilities to being less than perfect, we might not retain a single trace of information that is not considered necessary to the purpose of the communication.

The Moses illusion is the name for a common mistake that people make when asked a question such as, 'How many animals of each kind did Moses take on the ark?' (Erickson & Mattson, 1981, p. 542). The Moses illusion posits that when confronted with this question, they will answer 'two', failing to account for the fact that it was Noah who embarked with the animals on his ark and not Moses. This error shows that, though people know the facts (e.g., researchers verified that



participants knew that Noah went on the arc and not Moses), they do not scrutinize sentences to a degree where they detect that the information is inaccurate. When subjects are misdirected to pay attention to information that is incorrect, they are less likely to make the mistake than when it is the focal point of the sentence (Bredart & Modolo, 1988). This misparse is hypothesized to be caused by people only needing a partial match of a sentence in order to process it. It is both more cognitively efficient and useful in the real world to match imperfect but similar concepts in order to comprehend language. Indeed, research on the Moses illusion shows that the misinformation needs to be semantically related to the actual target in order for the error to be made (i.e., Moses has many broad features that are similar to Noah, such as they are both characters from the Bible; Park & Reder, 2004).

Importantly, this indicates that when taking part in a conversation and processing others' utterances, we might be inferring much more information than we are aware of. In an attempt to expend as little effort as possible, we base our listening on our expectations (Reder & Cleeremans, 1990). This is in accord with psycholinguistics research on sentence processing that shows that we typically only encode gist meaning. Indeed, we usually do not remember questions or statements word for word, but only the gist of what is being said (Sachs, 1967). Once a sentence is parsed and understood, we discard the exact phrasing (Bransford & Franks, 1971; Jarvella, 1971). Some research has shown evidence of encoding of surface form (Hornby, 1974), though it is typically thought to be remembered only for sentences that are distinctive (Kintsch & Bates, 1977), interactive (Keenan, MacWhinney, & Mayhew, 1977) and where the referent (i.e., what the speaker is referring to) is difficult to identify (Fletcher, 1984; Givón, 1983).

Baker and Wagner (1987) also examined the effect of false information by investigating where individuals focus in processing sentences. The researchers had participants evaluate the truthfulness of statements of facts that were assessed as common knowledge about science, history, famous people and places. False information was provided in either a subordinate clause (e.g., 'The liver,

which is an organ found only in humans, is often damaged by heavy drinking') or the main clause of the sentences (e.g., 'The liver, which is often damaged by heavy drinking, is an organ found only in humans'; p. 249). Participants were less accurate in their assessment of what was true and false when false information was in the subordinate clause. This indicates that people (e.g., interviewees being questioned) will likely do a less than perfect job at evaluating subordinate clauses of questions or information that they do not consider key to persuading the interviewer of their innocence.

Additionally, processing questions is cognitively effortful as it places demands on working memory (Buttner, 2007; Graesser et al., 1994). Questions contain unintegrated syntactic information called fillers, such as 'who' or 'why' that indicate that a sentence is a question and that hold the place of what the answer should be. When processing a question, one needs to hold this uncertain feature in mind while processing the rest of the sentence (Fanselow, Kliegl, & Schlesewsky, 1999; Fiebach, Schlesewsky, & Friederici, 2001). When interpreting any clause, people typically use the 'given-new strategy' (Clark & Haviland, 1974; 1977), where the presupposition would typically be the 'given' and the focal point would be the 'new' (Chomsky, 1971). The listener distinguishes new and given information (Chafe, 1970), identifies what information is known by conducting a memory search to identify such information and then integrates the new information at the place indicated by the memory search. This might be why presuppositions are given less attention, as their purpose is only to locate where to integrate the new information into the memory structure. The given-new strategy only functions efficiently if the listener is able to identify the information in his or her memory. If not, they must create a new memory structure to accommodate the new information and the sentence takes longer to interpret (Clark & Haviland, 1974). This is relevant in interview settings as the interviewee might not always be aware of the information referred to in the questions that they are asked. If this is the case, it would be harder for them to process the questions they are being asked.



Understanding how questions are encoded, processed, and stored in memory can be useful for interviewing research as it shows what information people will regard and critically evaluate and what will be encoded loosely and potentially more subject to biases and personal interpretation. Furthermore, it also provides an explanation of how information is integrated with knowledge that is already known, such as common ground or, in the case of a suspect interview, facts about past activities that may be related to the crime that the interview concerns.

Questions in Interview Settings

Research on questions in interview settings has focused mainly on the manner of asking questions and their content, with particular concern for suggestive questions. As will be discussed below, however, there is no exact definition of a suggestive question. Even apparently negligible elements of questions can imply facts, both true and untrue. Thus, although research has shown that suggestive questions can influence people's knowledge, this review addresses questions in general. The research discussed below is specific to leading questions, but provides support for the importance of question processing and hints at where people retrieve information from when processing questions in general.

With their exploration of suggestive questions, researchers have shown that the presupposition of a question can affect people's memory (Hornby, 1972; Loftus, 1975; Loftus, Miller, & Burns, 1978). For example, Harris (1973) found that framing or wording questions in different ways changed people's answers to them. For instance, asking subjects how tall or how short a basketball player was changed how participants responded. When asked, 'how short is the basketball player?' the average response was significantly lower than when asked to assess how tall he was (Harris, 1973). Even subtle changes in language can lead to different inferences (Bolinger, 1974). More specifically for conversational implicatures, Loftus and Zanni (1975) found that changing the article in a question from 'the' to 'a' increased participants' responses that they did not know whether they had seen something in a video. Indeed, using 'the' implies that the interlocutor knows of the existence

of the topic asked about (Anderson & Bower, 1973). Thus even such a small change in the phrasing of a question can have an effect on people's interpretation of it.

Memory researchers have also found evidence for a so-called misinformation effect (Loftus, 1975). In a first investigation of this effect, Loftus (1975) found that if a listener was unsure about the veracity of the presupposition of a question, they typically integrated it as true into their memory. Indeed, in five studies, Loftus demonstrated the effect of questions' presuppositions on subjects' memories of events. She found that asking participants to respond to questions about videos they had viewed could influence their later recall of those videos. For example, participants viewed a video of a car accident and were then asked a question about the speed of the car in the video. In one condition, the question included an allusion to a barn (i.e., 'How fast was the white sports car going when it passed the barn while traveling along the country road?' Loftus, 1975, p. 566). In the other condition there was no mention of the barn. In fact, there was no barn in the video and that presupposition of the question was false. Yet this falsified presupposition still influenced people's memory, making them think they saw a barn in the video. These studies supported the 'construction theory' of memory (Loftus, 1975, p. 564) that explains that when we remember an event, we reconstruct the event in our mind (see also, Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1973). Thus when we are asked a question about said event, the presupposed information, distorted or not, is incorporated into the memory representation and is then assimilated as such (Loftus et al., 1978).

Similarly, Dodd and Bradshaw (1980) found that the source of the questions (the person asking the questions) could attenuate the effects of false presuppositions on subjects' memories for events. In one study, using a similar paradigm as Loftus (1975) participants were either told that the questions were asked by the lawyer representing the person causing the car accident depicted in the video they watched (which would give them a reason to deceive) –or they received no such information. In a second study, the questions were attributed to either an innocent bystander (no



reason to deceive) or the driver of a car (reason to deceive). Participants who were told about the source of the questions were significantly less affected by the suggestive questions than those that were not or those who were told the source was someone they could trust (Dodd & Bradshaw, 1980). This is especially relevant here because interviewees might not trust their interviewer based on previous experience or general knowledge about the agency the interviewer represents. For example, a suspect might not trust a law enforcement officer or a foreign informant might not trust an intelligence officer representing a country other than their own. This means that people might naturally process their interviewer's statements with more scrutiny. Thus suspects might not be as likely to take the interviewer's questions at face value and might encode the information more critically according to their own beliefs.

Although the misinformation effect pertains to how people can unknowingly be affected by untrue information that can be included in questions, this review is more concerned with the information that is contained in the questions' presupposition and how people assimilate it into their memories, both purposefully and on an unconscious level. Indeed, the information that is asked about might not be untrue, but it could be unknown to the person being questioned, if they are innocent, for example. Regardless of guilt or innocence, a suspect may assimilate information from questions into their memory, which could affect their future responses. Given the level of imperfection with which people process language, it seems that question processing should be considered as an important factor in the research on the effects of questions during an interview.

Conclusions

This review summarized the literature on suspect strategies and an empirically supported interviewing strategy (i.e., the SUE technique); the theory of which we sought to build upon. Furthermore, this document surveyed several general psychological topics that

could be applied to strengthen both the utility and theoretical backbone of interviewing. As evidenced by the review, there is little overlap between basic and applied research on the effects of questions on suspects' perceptions of the evidence. However, because of their situation (i.e., being in a custodial interview, or even being suspicious of their interviewer) a suspect should be motivated to evaluate the interviewer's point of view and strategically draw inferences from their questions. Research on interviewing techniques and suspect thought processes could benefit from incorporating findings from basic psycholinguistic and perspective taking research. From the literature reviewed, we can draw several conclusions about how to study the inferences that suspects draw from the questions they are asked during an interview.

The perspective taking research informs us as to what suspects' thought processes would resemble when attempting to read the interviewer's mind: From the results presented we see that it might be quite hard for a suspect to discount their own perspective when inferring what the information the interviewer holds against them. This means that in a SUE type interview, when asked about specific pieces of information, the suspect's own knowledge might cloud his or her judgment about what information the interviewer has. Further, this egocentric bias might be enhanced or reduced under certain circumstances (e.g., different levels of cognitive load, deceptiveness, etc.) when suspects are better able to take the interviewer's perspective. Indeed, findings from research on perspective taking and common ground suggest that suspects under cognitive load, or those not motivated to take the perspective of the interviewer might perform more egocentrically and make assessments of the interviewer's knowledge disproportionately based on their own knowledge. This is strengthened by the linguistic research supporting the idea that we only encode gist information. If we only remember the gist of what someone asked and we know information that is similar to what the interviewer is asking about, source mon-



itoring errors could happen and the suspect could think that the interviewer knows information because the suspect knows it themselves. Thus future research should specifically focus on investigating the source of the information suspects are using as a result of their attempted perspective taking. Additionally, based on the psycholinguistics research overviewed, one could hypothesize that suspects will be more likely to remember information from the focus of the questions and that their inferences about information contained in the presuppositions and subordinate clauses will be less accurate than those based on the focus. Future research could examine the role of question complexity on inferences that suspects draw from the questions. Bridging the three general topics discussed, we hypothesize that the more complex the question, the more egocentric the suspect's perspective taking efforts would be. Hence future research should explore the compounded effect of different types of questions (e.g., with or without subordinate clauses) and perspective taking on suspect's inferences about interviewers' knowledge of evidence.

In conclusion, we suggest that to increase the theoretical soundness of interviewing techniques, future research should concentrate on inserting the basic principles outlined in this review into interviewing research. To do this, we propose a model of factors that can influence the suspect's decision making process (see Figure 1): When asked a question, a suspect processes it through the lens of his or her prior knowledge and encodes the information through inferences based on common ground knowledge shared by them and the interviewer. When attempting to take the interviewer's perspective, the suspect will base their judgment of what evidence they think the interviewer has on what information they encoded. Perspective taking activities will also be affected by inferences about common ground knowledge and biased by their own prior knowledge. The process is cyclical, where previous questions, inferences about these questions, and previous perspective taking will influence suspects' interpretation of questions

going forward during an interview. Examining the specific effect of these influential factors on suspects' perceptions of evidence held of an interviewer may deepen the understanding of the complexities of suspects thought processes during an interview. Using the information processing model described above in the context of interviews might help provide more clues to distinguish between guilty and innocent suspects and help understand what effects the questions an interviewer asks can have on a suspect's perception of the evidence, but also how these effects can be used to craft more effective interview techniques, based on psychological theory.

This review of the literature takes a cognitive approach to the interview process whereby the concern is applying psychological theory to assess suspects' cognitive ability to perceive (accurately) the evidence that an interviewer has against them. The interviewing research on differences between innocent and guilty suspects suggests that guilt should be a motivating factor in increasing the likelihood that a suspect would be motivated to draw inferences from the questions an interviewer asks. We thus assumed that guilty suspects would have more motivation to engage in instrumental mind reading. However, the power of a situation should not be ignored, as other situational variables beyond guilt (e.g., suspicion on the part of the suspect, accusation from the interviewer) could create the motivation for any interviewee to engage, or at least attempt to engage in the activities described in this review.



References

- Abel, T. (1948). The operation called Verstehen. *American Journal of Sociology*, 54, 211-218.
- Afifi, W. A., & Weiner, J. L. (2004). Toward a theory of motivated information management. *Communication Theory*, 14, 167-190. doi: 10.1111/j.1468-2885.2004.tb00310.x
- Afifi, W. A., & Weiner, J. L. (2006). Seeking information about sexual health: Applying the theory of motivated information management. *Human Communication Research*, 32, 35-57. doi: 10.1111/j.1468-2958.2006.00002.x
- Alicke, M. D., Dunning, D., & Krueger, J. I. (Eds.). (2005). *The self in social judgment*. New York, NY: Taylor & Francis Group.
- Allport, F. H. (1924). The group fallacy in relation to social science. *American Journal of Sociology*, 29, 688-706.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256-274. doi: 10.1037/0033-2909.111.2.256
- Ames, D. R. (2004a). Inside the mind-reader's toolkit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, 87, 340-353. doi: 10.1037/0022-3514.87.3.340
- Ames, D. R. (2004b). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, 87, 573-585. doi: 10.1037/0022-3514.87.5.573
- Ames, D. L., Jenkins, A. C., Banaji, M. R., & Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychological Science*, 19, 642-644. doi: 10.1111/j.1467-9280.2008.02135.x
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, D.C: VH Winston & Sons.
- Apperly, I A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, 17, 841-844. doi: 10.1111/j.1467-9280.2006.01791.x
- Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, 45, 190-201. doi: 10.1037/a0014098
- Astington, J. W., Harris, P. L., & Olson, D. (Eds.). (1988). *Developing theories of mind*. Cambridge, England: Cambridge University Press.
- Baker, L., & Wagner, J. L. (1987). Evaluating information for truthfulness: The effects of logical subordination. *Memory & Cognition*, 15, 247-255. doi: 10.3758/BF03197723
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, 109, 18-40. doi: 10.1016/j.cognition.2008.07.005
- Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 901-938). Amsterdam, Netherlands: Elsevier.



- Belnap, N. D. (1966). Questions, answers, and presuppositions. *The Journal of Philosophy*, 63, 609-611.
- Bierhoff, H. W. (1989). *Person perception and attribution*. Berlin: Springer-Verlag.
- Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18, 382-386. doi: 10.1111/j.1467-9280.2007.01909.x
- Bless, H., & Igou, E. R. (2005). Mood and the use of general knowledge structures in judgment and decision making. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 193-210). San Diego, CA: Academic.
- Bolinger, D. (1974). *Meaning and form*. *Transactions of the New York Academy of Sciences*, 36, 218-233. doi: 10.1111/j.2164-0947.1974.tb01567.x
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3, 193-209. doi: 10.1016/0010-0285(72)90003-5
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331-350. doi: 10.1016/0010-0285(71)90019-3
- Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Brédart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67, 135-144. doi: 10.1016/0001-6918(88)90009-1
- Brimbal, L, Hartwig, M., & Luke, T. J. (2013). [Suspects' pragmatic inferences about evidence during a short interview]. Unpublished raw data.
- Brown, P. M., & Dell, G. S. (1987) Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19, 441-72. doi: 10.1016/0010-0285(87)90015-6
- Buttner, A. C. (2007). Questions versus statements: Challenging an assumption about semantic illusions. *The Quarterly Journal of Experimental Psychology*, 60, 779-789. doi: 10.1080/17470210701228744
- Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning and Individual Differences*, 15, 53-65. doi: 10.1016/j.lindif.2004.07.002
- Chandler, M. (1976). Social cognition: A selective review of current research. In W. F. Overton & J. M. Gallagher (Eds.), *Knowledge and development*. New York: Plenum Press
- Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In L. A. Jakobovits & D. Steinberg (Eds.), *Semantics: An interdisciplinary reading in philosophy, psychology, linguistics, and anthropology* (pp. 183-216). Cambridge, England: Cambridge University Press.
- Cirilo, R. K. (1981). Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, 20, 358-367. doi: 10.1016/S0022-5371(81)90505-3
- Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance* (pp. 313-330). Hillsdale, N.J.: Erlbaum.



- Clark, H. H., & Haviland, S. E. (1974). Psychological processes as linguistic explanation. In D. Cohen (Ed.), *Explaining linguistic phenomena* (pp. 91-124). Washington, D.C.: Hemisphere Publishing Corporation.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshe, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-61). Cambridge: Cambridge University Press.
- Clark, H. H., & Schaeffer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259-294. doi: 10.1207/s15516709cog1302_7
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology*, 36, 752-766. doi: 10.1037/0022-3514.36.7.752
- Colwell, K., Hiscock-Anisman, C., Memon, A., Woods, D., & Michlik, P. M. (2006). Strategies of impression management among deceivers and truth-tellers: How liars attempt to convince. *American Journal of Forensic Psychology*, 24, 31-38.
- Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects Theory-of-Mind use. *Emotion*, 8, 725-730. doi: 10.1037/a0013283
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84-107.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126. doi: 10.1037/0022-3514.44.1.113
- Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes*, 65, 201-211. doi: 10.1006/obhd.1996.0020
- DeCoster, J., & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review*, 8, 2-27. doi: 10.1207/S15327957PSPR0801_1
- Dell, G.S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener." In D. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics* (pp. 105-129). New York: Academic Press.
- Dillon, J. T. (1982). The multidisciplinary study of questioning. *Journal of Educational Psychology*, 74, 147-165. doi: 10.1002/j.2162-6057.1982.tb00326.x
- Dixon, P., LeFevre, J. A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, 80, 465-472. doi: 10.1037/0022-0663.80.4.465
- Dobson, M., & Markham, R. (1993). Imagery ability and source monitoring: Implications for eyewitness memory. *British Journal of Psychology*, 84, 111-118.



- Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, 19, 695-704. doi: 10.1016/S0022-5371(80)90379-5
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436.
- Epley, N. & Caruso, E. M. (2008). Perspective taking: Misstepping into others' shoes. In K. D. Markam, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 297-311). New York, NY: Taylor & Francis Group.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327-339. doi: 10.1037/0022-3514.87.3.327
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40, 760-768. doi: 10.1016/j.jesp.2004.02.002
- Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gardner, & G. Lindzey (Eds.), *Handbook of social psychology: Volume one* (pp. 498-541) doi: 10.1002/9780470561119.socpsy001014
- Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540-551. doi: 10.1016/S0022-5371(81)90165-1
- Evans, J. R., Meissner, C. A., Brandon, S. E., Russano, M. B., & Kleinman, S. M. (2010). Criminal versus HUMINT interrogations: The importance of psychological science to improving interrogative practice. *Journal of Psychiatry & Law*, 38, 215-249.
- Eysenck, M., & Keane, M. T. (2010). *Cognitive psychology: A student's handbook, 6th edition*. New York, NY: Psychology Press.
- Fahsing, I. A., & Rachlew, A. (2009). Investigative interviewing in the Nordic region. In T. Williamson, B. Milne, & S. Savage (Eds.), *International developments in investigative interviewing* (pp. 39-65). Cullompton: Willan Publishing.
- Fanselow, G., Kliegl, R., & Schlesewsky, M. (2005). Syntactic variation in German wh-questions: a rating and a training study. *Linguistic Variation Yearbook*, 5, 37-63. doi: 10.1075/livy.5.03fan
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40, 296-340. doi: 10.1006/cogp.1999.0730
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, 30, 321 - 338. doi: 10.1023/A:1010447102554
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1-74. doi: 10.1016/S0065-2601(08)60317-2
- Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture 2nd edition*. Thousand Oaks, CA: SAGE Publications, Inc.



- Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist*, 41, 418–425. doi: 10.1037/0003-066X.41.4.418
- Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, 24, 15-23 doi: 10.1080/016502500383421
- Fletcher, G. J. (1984). *Psychology and common sense*. *American Psychologist*, 39, 203-213. doi: 10.1037/0003-066X.39.3.203
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117, 39–66. doi: 10.1037/0033-2909.117.1.39
- Foss, C. C., & Bower, G. H. (1986). Understanding actions in relation to goals. In N. E. Sharkey (Ed.), *Advances in cognitive science 1* (pp. 94-124). Chichester: Ellis Horwood.
- Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, 19, 378-384. doi: 10.1111/j.1467-9280.2008.02096.x
- Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, 78, 211–222. doi: 10.1037/0022-3514.78.2.211
- Gilovich, T., Savitsky, K., & Medvec, V.H. (1998). The illusion of transparency: Biased assessments of others' ability read one's emotional states. *Journal of Personality and Social Psychology*, 75, 332-346. doi: 10.1037/0022-3514.75.2.332
- Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study*. doi: 10.1075/tsl.3
- Goetz, E. T. (1979). Inferring from text: Some factors influencing which inferences will be made. *Discourse Processes*, 2, 179-195. doi: 10.1080/01638537909544464
- Goffman, E. (1974). *Frame analysis: An essay on the organization experience*. Boston: Northeastern University Press.
- Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7, 145–171. doi: 10.1111/j.1468-0017.1992.tb00202.x
- Gopnik, A., & Wellman, H. M. (1994). The “theory” theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
- Gopnik, A. (1993). Theories and illusions. *Behavioral and Brain Sciences*, 16, 90-100. doi: 10.1017/S0140525X00029253
- Graesser, A. C., McMahan, C. L., & Johnson, B. K. (1994). Question asking and answering. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 517–538). New York: Academic Press.
- Granhag, P. A., Clemens, F., & Strömwall, L. A. (2009). The usual and the unusual suspects: Level of suspicion and counter-interrogation tactics. *Journal of Investigative Psychology and Offender Profiling*, 6, 129-137. doi: 10.1002/jip.101
- Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, 14, 189-200. doi:



10.1080/10683160701645181

- Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence (SUE) technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception Detection: Current Challenges and new approaches* (pp. 231–251). New York, NY: John Wiley & Sons.
- Granhag, P. A., Cancino Montecinos, S., & Oleszkiewicz, S. (2013). Eliciting intelligence from sources: The first scientific test of the Scharff technique. *Legal and Criminological Psychology*. Advance online publication. doi: 10.1111/lcrp.12015
- Granhag, P. A., Strömwall, L. A., Willén, R. M., & Hartwig, M. (2013). Eliciting cues to deception by tactical disclosure of evidence: The first test of the evidence framing matrix. *Legal and Criminological Psychology*, 18, 341–355. doi: 10.1111/j.2044-8333.2012.02047.x
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27. doi: 10.1037/0033-295X.102.1.4
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds), *Syntax and Semantics III: Speech Acts* (pp. 41–58). New York, NY: Academic Press.
- Griffiths, A., & Milne, R. (2006). Will it all end in tiers? Police interviews with suspects in Britain. In T. Williamson (Ed.), *Investigative Interviewing: Rights, Research, Regulation*, (pp. 167–189). Portland: Willan Publishing.
- Hafer, C. L., & Bègue, L. (2005). Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin*, 131, 128–167. doi: 10.1037/0033-2909.131.1.128
- Hanna, J., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28, 105–115. doi: 10.1016/j.cogsci.2003.10.002
- Hanna, J., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6
- Harris, R. J. (1973). Answering questions containing marked and unmarked adjectives and adverbs. *Journal of Experimental Psychology*, 97, 399–401. doi: 10.1037/h0034165
- Hartwig, M., Granhag, P. A., & Luke, T. J. (2014). Strategic use of evidence during investigative interviews: The state of the science. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 3–38). Oxford: Elsevier.
- Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, 13, 213–227. doi: 10.1080/10683160600750264
- Hartwig, M., Granhag, P. A., Strömwall, L., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects [Special issue]. *The Open Criminology Journal*, 3, 10–16.
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews. *Law and Human Behavior*, 30, 603–619. doi: 10.1007/s10979-006-9053-9
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic



- disclosure of evidence. *Law and Human Behavior*, 29, 469-484. doi: 10.1007/s10979-005-5521-x
- Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging*, 5, 119-126. doi: 10.1037/0882-7974.5.1.119
- Henkel, L. A. & Coffman, K. J. (2004). Memory Distortions in Coerced False Confessions: A Source Monitoring Framework Analysis. *Applied Cognitive Psychology*, 18, 567-588. doi: 10.1002/acp.1026
- Higgins, E. T. (1981). Role-taking and social judgment: alternative developmental perspectives and processes. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 119-153). Cambridge: Cambridge University Press.
- Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal & Criminological Psychology*, 13, 357-371. doi: 10.1348/135532507X238682
- Hines, A., Colwell, K., Hiscock-Anisman, C., Garrett, E., Ansarra, R., & Montalvo, L. (2010). Impression management strategies of deceivers and honest reporters in an investigative interview. *The European Journal of Psychology Applied to Legal Context*, 2, 73-90.
- Hornby, P. A. (1972). The psychological subject and predicate. *Cognitive Psychology*, 3, 632-642. doi: 10.1016/0010-0285(72)90023-0
- Hornby, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior*, 13, 530-538. doi: 10.1016/S0022-5371(74)80005-8
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96, 127-142. doi: 10.1016/j.cognition.2004.07.001
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91-117. doi: 10.1016/0010-0277(96)81418-1
- Houston, K. A., Meissner, C. A., & Evans, J. R. (2014). Psychological processes underlying true and false confessions. In Bull, R. (Ed.), *Investigative interviewing* (pp. 19-34). New York: Springer.
- Inbau, F., Reid, J., Buckley, J., & Jayne, B. (2011). *Criminal interrogation and confessions*. Burlington, MA: Jones & Bartlett Learning.
- Isen, A. M. (1984). Towards understanding the role of affect in cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 179-236). Hillsdale, NJ: Erlbaum.
- Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 10, 409-416. doi: 10.1016/S0022-5371(71)80040-3
- Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology*, 117, 390-394. doi: 10.1037/0096-3445.117.4.390
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28. doi: 10.1037/0033-2909.114.1.3
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85. doi: 10.1037/0033-295X.88.1.67



- Karniol, R. (1990). Reading people's minds: A transformation rule model for predicting others' thoughts and feelings. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. San Diego, CA: Academic Press, Inc.
- Karniol, R. (2003). Egocentrism versus protocentrism: The status of self in social prediction. *Psychological Review*, 110, 564-580. doi: 10.1037/0033-295X.110.3.564
- Kass, R., & Finin, T. (1988). Modeling the user in natural language systems. *Computational Linguistics*, 14, 5-22.
- Kassin, S.M. (2005). On the psychology of confessions: Does innocence put innocent at risk? *American Psychologist*, 60, 215-228. doi: 10.1037/0003-066X.60.3.215
- Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A. M., & La Fon, D. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior*, 31, 381-400. doi: 10.1007/s10979-006-9073-5
- Katz, D., & Allport, F. H. (1931). *Students' Attitudes: A Report of the Syracuse University Reaction Study*. Syracuse, NY: Craftsman Press.
- Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, 16, 549-560. doi: 10.1016/S0022-5371(77)80018-2
- Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, 26, 165-208. doi: 10.1006/cogp.1994.1006
- Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they "should". In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150-166). Cambridge, U.K: Cambridge University Press.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32-38. doi: 10.1111/1467-9280.00211
- Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39, 1-20. doi: 10.1006/jmla.1998.2563
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25-41. doi: 10.1016/S0010-0277(03)00064-7
- King, L., & Snook, B. (2009). Peering inside a Canadian interrogation room: An examination of the Reid model of interrogation, influence tactics, and coercive strategies. *Criminal Justice & Behavior*, 36, 674-694. doi: 10.1177/0093854809335142
- Kintsch, W., & Bates, E. (1977) Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 150-159. doi: 10.1037/0278-7393.3.2.150
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394. doi: 10.1037/0033-295X.85.5.363
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4, 343-346.



doi: 10.1037/h0023705

- Krueger, J. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social perception* (pp. 17-41). New York, NY: Taylor & Francis Group
- Krueger, J., & Clement, R. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, 65, 596-610. doi: 10.1037/0022-3514.67.4.596
- Lerner, M.J. (1980). *The belief in a just world*. New York: Plenum.
- Lerner, M. J. & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, 85, 1030-1051. doi: 10.1037/0033-2909.85.5.1030
- Lerner, M. J., & Simmons, C. H. (1966). Observer's reaction to the "innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology*, 4, 203-210. doi: 10.1037/h0023562
- Lewis, C., & Mitchell, P. (Eds.). (1994). *Children's early understanding of mind: Origins and development*. Hove, England: Erlbaum.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46, 551-556. doi: 10.1016/j.jesp.2009.12.019
- Locke, E. (2000). Motivation, cognition, and action: An analysis of studies of task goals and knowledge. *Applied Psychology*, 49, 408-429. doi: 10.1111/1464-0597.00023
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560-572. doi: 10.1016/0010-0285(75)90023-7
- Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, 5, 86-88. doi: 10.3758/BF03336715
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31. doi: 10.1037/0278-7393.4.1.19
- Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language*, 31, 634-647. doi: 10.1016/0749-596X(92)90032-S
- Luke, T. J., Dawson, E., Hartwig, M., & Granhag, P. A. (2014a). How awareness of evidence induces forthcoming counter-interrogation strategies. *Applied Cognitive Psychology*. Advance online publication. doi: 10.1002/acp.3019
- Luke, T. J., Hartwig, M., Shamash, B., & Granhag, P.A. (March, 2014b). Suspect counter-interrogation strategies: The effect of prior knowledge of the interviewing techniques. Paper presented at meeting of the American Psychology-Law Society, New Orleans, LA.
- Maccoby, E. E. (1980). *Social development*. New York: Harcourt Brace Jovanovich.
- Malle, B. F. (2005). Folk theory of mind: Conceptual foundations of human social cognition. In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 225 - 255). New York: Oxford University Press.



- Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, 75, 314-318. doi: 10.1037/0022-0663.75.2.314
- May, L., Granhag, P. A., & Oleszkiewicz, S. (2014) Eliciting Intelligence Using the Scharff-Technique: Closing in on the Confirmation/Disconfirmation-Tactic. *Journal of Investigative Psychology and Offender Profiling*, 11, 136-150. doi: 10.1002/jip.1412
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287-330. doi: 10.1037/0033-295X.86.4.287
- McCornack, S. A. (1992). Information Manipulation Theory. *Communication Monographs*, 59, 1-16. doi: 10.1080/03637759209376245
- Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*. Chichester: Wiley.
- Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 179-195). Oxford, NY: Oxford University Press, Inc.
- Mitchell, K. J. & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, 135, 638-677. doi: 10.1037/a0015849
- Mitchell, P., & Riggs, K. J. (Eds.). (2000). *Children's reasoning and the mind*. Hove, England: Psychology Press.
- Moore, D.A. (2005). Myopic biases in strategic social prediction: Why deadlines put everyone under more pressure than everyone else. *Personality and Social Psychology Bulletin*, 31, 668-679. doi: 10.1177/0146167204271569
- Nadig, J. S., & Sedivy, J. C. (2002) Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329-36. doi: 10.1111/j.0956-7976.2002.00460.x
- Narchet, F. M., Meissner, C. A., & Russano, M. B. (2011). Modeling the influence of investigator bias on the elicitation of true and false confessions. *Law and Human Behavior*, 35, 452-465. doi: 10.1007/s10979-010-9257-x
- Neale, M. A., & Bazerman, M. H. (1983). The role of perspective-taking ability in negotiating under different forms of arbitration. *Industrial and Labor Relations Review*, 36, 378-388.
- Nickerson, R. S. (1999). How we know—And sometimes misjudge—What others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737-759. doi: 10.1037/0033-2909.125.6.737
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259. doi: 10.1037/0033-295X.84.3.231
- Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 476-492. doi: 10.1037/0278-7393.22.2.476
- Oleszkiewicz, S., Granhag, P. A., & Cancino Montecinos, S. (2014). The Scharff-technique: Eliciting intelligence from human sources. *Law and Human Behavior*. Advance online publication. doi:



10.1037/lhb0000085

- Park, H., & Reder, L. M. (2004). Moses illusion: Implication for human cognition. In Pohl, R. F. (Ed.), *Cognitive illusions* (pp. 275-291). Hove: Psychology Press.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J., & Kühberger, A. (2005). Mental simulation: Royal road to other minds? In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 174-189). New York, NY: The Guilford Press.
- Perner, J., Ruffman, T., & Leekam, S.R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development*, 65, 1228-1238. doi: 10.1111/j.1467-8624.1994.tb00814.x
- Phillips, L. H., Bull, R., Adams, E., & Fraser, L. (2002). Positive mood and executive function: Evidence from Stroop and fluency tasks. *Emotion*, 2, 12-22. doi: 10.1037/1528-3542.2.1.12
- Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press. (Original work published 1932)
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 1-22. doi: 10.1017/S0140525X04000056
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526. doi: 10.1017/S0140525X00076512
- Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: the Moses illusion revisited. *Psychology of Learning and Motivation*, 25, 233-258. doi: 10.1016/S0079-7421(08)60258-3
- Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279-301. doi: 10.1016/0022-1031(77)90049-X
- Rosnagel, C. (2000). Cognitive load and perspective taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30, 429-445. doi: 10.1002/(SICI)1099-0992(200005)
- Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science*, 16, 481-486. doi: 10.1111/j.0956-7976.2005.01560.x
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442. doi: 10.3758/BF03208784
- Sarbin, T. (1954). Role theory. In G. Lindzey (Ed.), *Handbook of social psychology* (pp. 223-258). Cambridge, MA: Addison-Wesley.
- Saxe, R. (2008). The happiness of the fish: Evidence for a common theory of one's own and others' actions. In K. D. Markam, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 257-266). New York, NY: Taylor & Francis Group.
- Schank, R. (1982). *Dynamic memory: A theory of learning in computers and people*. London: Cambridge University Press.



- Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327. doi: 10.1515/semi.1973.8.4.289
- Singer, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479–515). San Diego, CA: Academic Press.
- Snook, B. & Barron, T. (2013). An alternative to deceptive police interrogation tactics. *Legal Aid Society*. Lecture conducted from St. Francis College, Brooklyn, NY.
- Stalnaker, R.C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics, Vol. 9: Pragmatics* (pp. 315–332). New York: Academic Press
- Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, 25, 701–721. doi: 10.1023/A:1020867916902
- Strack, F., Schwartz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, 9, 111–125. doi: 10.1521/soco.1991.9.1.111
- Strawson, P. F. (1952). *Introduction to logical theory*. New York: Wiley & Sons.
- Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, 12, 207–219. doi: 10.1080/10683160512331331328
- Tanenhaus, M. K., Spivey, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. doi: 10.1126/science.7777863
- Thompson, L., & Hastie, R. (1990). Social perception in negotiation. *Organizational Behavior and Human Decision Processes*, 47, 98–123. doi: 10.1016/0749-5978(90)90048-E
- Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review*, 5, 449–453. doi: 10.3758/BF03208820
- Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, 29, 1159–1168. doi: 10.1177/0146167203254597
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10, 141–142. doi: 10.1016/j.tics.2006.02.003
- Vrij, A., Granhag, P. A., & Mann, S. (2010). Good liars. *Journal of Psychiatry & Law*, 38, 77–98.
- Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, 20, 28–32. doi: 10.1177/0963721410391245
- Vrij, A., Mann, S., & Fisher, R. P. (2006). An empirical test of the Behaviour Analysis Interview. *Law and Human Behavior*, 30, 329–345. doi: 10.1007/s10979-006-9014-3
- Vrij, A., Mann, S., & Leal, S. (2013). Deception traits in psychological interviewing. *Journal of Police and Criminal Psychology*, 28, 115–126. doi: 10.1007/s11896-013-9125-y
- Walker, C. H., & Meyer, B. J. (1980). Integrating different types of information in text. *Journal of*



- Verbal Learning and Verbal Behavior*, 19, 263-275. doi: 10.1016/S0022-5371(80)90221-2
- Wegner, D. M. (2005). Who is the controller of controlled processes? In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 19–36). New York: Oxford University Press.
- Wegner, D. M., Wenzlaff, R., Kerker, R. M., & Beattie, A. E. (1981). Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology*, 40, 822-832. doi: 10.1037/0022-3514.40.5.822
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.
- Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, 14, 133-145. doi: 10.1080/01638539109544779
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441-517. doi: 10.1006/jmla.2002.2864
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, 14, 1-59. doi: 10.1016/S0065-2601(08)60369-X



Assessing Polygraph Accuracy; The Importance of Choosing an Evaluation Technique Which Is Compatible With the Way the Examinations Were Conducted

Avital Ginton

AV-GN Polygraph, Applications & Research, Israel

Abstract

Sterzer and Elaad in their Israel National Police polygraph field study on Comparison Question Test (Sterzer & Elaad, 1985), found that the false positive (FP) error rate in severe crimes was almost double compared to minor crimes. This was attributed to the difference in the degree of threat and the perceived consequences in the two levels of crime severity. Another factor, however, might have contributed to the difference in FP error rates found in that study. During the year 1979, the Israel National Police Polygraph Laboratory went through a substantial change in the way the polygraph examinations were conducted and evaluated. Unfortunately, this was ignored in the study, and the current analysis indicates that a great portion the crime severity effect might be an artifact of this change. On a more general level that goes beyond that study, this case demonstrates the importance of choosing an evaluation technique which is compatible with the way the examinations were conducted.

Sterzer and Elaad in their study on Comparison Question Test (IDENTA, 1985) found that with numerical scoring technique, using a zero cutoff point with no inconclusive zone, results in false positive (FP) error rate of 23.33% for a minor crime sample (MC) and 43.14% for a severe crime sample (SC). The difference between these two error rates, which was found to be statistically significant, was attributed to the difference in the degree of threats posed by the perceived consequences of the relevant questions in the two levels of crime severity. Another factor, however, might have contributed to the difference in FP error rates.

During the year 1979, the Israel National Police Polygraph Laboratory went

through a substantial change about the way the polygraph examinations were conducted and evaluated. The basic change was moving from a considerable reliance on the behavioral symptoms of the examinee and global evaluation of the polygraph records (as recommended by certain schools of thought), towards a major reliance on the semi-objective numerical scoring technique, which had by then become widespread. It was soon found that the change in the way information from the polygraph examinations is considered and evaluated, brought about a change in the manner by which the examinations were conducted, including a tendency to stress, during the pre-test interview and between charts, the comparison (control) questions more than had been done previously. This tendency resulted

1 This short article is an elaboration of a comment made on a study presented in IDENTA 85 by Sterzer & Elaad: Validity of the control question test in two levels of the severity of crimes. The original study and the comment were published in *Anti-Terrorism; Forensic Science; Psychology in Police Investigations*. Proceeding of IDENTA 85; the international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel Boulder, Colo.: Westview Press; Jerusalem: Heiliger and Co., 1986.

2 At the time the original comment was made, Dr. Ginton was heading the Behavior Section in the Criminal Identification and Forensic Sciences Division of Israel National Police & a Faculty member in the Criminology Department, Bar-Ilan University, Ramat-Gan, Israel.



in considerable increased physiological reactions to those questions.

That was evident in the records of the innocent examinees from the SC sample. Eighteen examinations were conducted during 1977-8, before the change took place, and twenty-two after it, during 1980-84³. It was found that in the earlier group, 55.6% of the records received negative numerical scores, (i.e., the reactions to the relevant questions in those records were stronger than to the control ones) compared to only 31.8% in the later one. Using the normal approximation to the binomial distribution, it was found that the probability associated with this difference is equal to or less than 0.065, ($Z=1.487$, one tailed) which is very close to an acceptable level of statistical significance.

In contrast, the MC sample was taken mostly (27 of 30 records) from examinations conducted in the 1980s (1980-4), and the percentage of negative scores for these 27 records was 18.5%.

It seems that in order to obtain an estimate for the effect of crime severity on FP error rates in this study, only examinations conducted in the 1980s should be considered. When this is done it was found that the difference between the MC and SC samples regarding the FP error rates – 18.5% and 31.8% respectively (with no inconclusive zone), is associated with the probability of equal to or less than 0.142 ($Z=1.07$, one tailed), which is far from any acceptable level of significance. It means that, although a tendency towards the predicted direction of difference in FP error rates was observed, there is still a chance of 14% that the observed effect is due to sample error or chance fluctuation.

Furthermore, as correctly mentioned by Sterzer and Elaad, in actuality, a safeguard against FP (as well as FN) is taking place by using an inconclusive zone for low score outcomes. For instance, UTAH approach recommended using cut scores of ± 6 to make a

call, and records that receive scores between plus 5 and minus 5 are deemed inconclusive (Barland and Raskin, 1975; Raskin and Honts, 2002).

Applying this rule to the 1980s sub-samples of innocent examinees yields FP error rates of only 11.7% (two errors in 17 conclusive records) and 8% (two errors in 24 conclusive records) for the SC and MC samples respectively. Thus, it seems that practically, the factor of crime severity has at most a small effect upon the FP error rates, and in order to better explore its actual existence a much larger sample is needed.

The main reason for this comment goes beyond the study being discussed. It demonstrates that one could not get a fair estimation of the accuracy of polygraph examinations by using an evaluation technique at odds with the way in which the examinations were conducted. Examinations, whose conduct was geared towards the global evaluation approach which includes for instance also the identification of developing patterns and trends along the repeated charts and behavioral symptoms, might be unsuitable for analysis by the more objective and strict numerical scoring technique, or even by blind global evaluation, which is not accompanied by behavioral-symptoms analysis. Some support for this conclusion can be drawn from a validity study which was conducted in the Israeli Police in 1979 and published a bit later (Ginton, Daie, Elaad & Ben-Shakhar, 1982). In that study, the polygraph tests and their analysis took place during the beginning of the transient period mentioned above concerning the manner by which the examinations were conducted and evaluated. Results of that study indicated that using numerical scoring was somewhat inferior compared to the global evaluation which was more compatible with the way the tests were conducted. Unfortunately, the point of compatibility was not addressed in the original study. This notion of compatibility considerations might be relevant in quite a number of other situations. For instance, two charac-

3 The SC sample also included 11 tests of innocent examinees from 1979, the year in which the change took place.



teristics of the Utah technique are 1) to review and discuss in a certain manner the questions between repeated charts and 2) in the scoring stage of a single-issue sequence, the Relevant questions are compared only to their adjacent preceding Comparison question (Raskin & Kircher, 2014). Unlike that, the Federal ZCT recommends as a default: 1) not to discuss the questions between charts and 2) in scoring the test (You-Phase) to pick for comparison the strongest reaction from the two Comparison questions surrounding each Relevant question (Federal Examiner Handbook, 2011). Applying the Utah scoring technique in that respect, to a ZCT examination in which no inter charts discussion took place, might results

in an increased FP rate. In the same vein, applying the federal scoring method in that respect to an examination that was conducted with inter charts discussion might results in a FN increased rate. That is because the inter charts discussion in the manner suggested by Utah method tends to increase in particular the salience of the Comparison question.

Unfortunately, this compatibility issue has been, overlooked in many studies, leading to unreliable and conflicting results. Finally, it is recommended to assimilate and incorporate this insight not only into the research arena but also to the everyday practice.



References

- Barland, G.H. and Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12, 321-330.
- Federal Examiner Handbook, (2011). Federal psychophysiological detection of deception examiner handbook. Special Edition. *Polygraph*, 40, 1, 1-66.
- Ginton, A., Daie, N., Elaad, E. and Ben-Shakhar, G. (1982). A method for evaluating the use of polygraph in a real-life situation. *Journal of applied psychology*, 67, 131-137.
- Raskin, D.C. and Honts, C, (2002). The comparison question test, in: Kleiner, M. (ed.) *Handbook of Polygraph Testing*. (2002), pp. 1-48, Academic Press, San-Diego, California, USA.
- Raskin, D.C. and Kircher, J.C., (2014). Validity of polygraph techniques and decision methods. In: Raskin, D.C., Honts, R.C. and Kircher, J.C. (Eds). *Credibility Assessment Scientific Research and Applications*. (2014), Pp. 65-132. Elsevier, Academic Press, San Diego, CA, USA,
- Sterezer, G. and Elaad, E. (1985). Validity of the control question test in two levels of the severity of crimes. In: *Anti-Terrorism; Forensic Science; Psychology in Police Investigations*. Proceeding of IDENTA 85. The international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel. Boulder, Colo.: Westview Press; Jerusalem: Heiliger and Co., 1986,



The Use of Polygraph Testing for Theft Investigation in Private Sector Institutions

Laura Pérez-Campos Mayoral¹, Eduardo Pérez-Campos Mayoral^{* 1, 2}, Gabriel Mayoral Andrade³, Carlos Perezcampos Mayoral³, Rocío Martínez Helmes³, Eduardo Pérez-Campos^{1, 4}

Abstract

One recurring problem that every business faces is internal theft. In this work the authors review alternative solutions and propose possible ways to reduce employee theft and embezzlement and improve in-company security through the use of the polygraph test. The primary objective of this procedure is to reduce risk by hiring people with appropriate profiles through analysis of the information generated in the pretest phase of the polygraph test. The second objective is the verification of the information obtained, through specific questions and proper analysis of the psychophysiological data acquired.

Key words: polygraph, theft, business.

The effects of theft-related criminal activity are a burden on society. They can cause a perception of personal insecurity leading to a change in habits, e.g. avoiding the area where a crime has been committed to minimize the risk of becoming a victim. However, shoplifting committed by customers or employees normally goes unnoticed and it is considered an “unaccountable loss” within budgets. Officially, theft is classified within ‘crimes against property’. Based on statistics from the Chamber of Commerce in the United States of America, employee theft costs companies over \$ 40 billion per year (Beedle, 2005).

This study highlights the importance of the use of polygraph testing within private institutions, more specifically, for its value as a diagnostic and preventive tool. We think the polygraph is a good investment for financial security and business management when used as a preventative warning to discourage losses due to misconduct or theft. Loss prevention in regard to theft is becoming increasingly important in strategic company planning, especially in the retail sector where loss from “shrinkage” is growing. Various studies conducted by the European Union reveal that theft in commercial distribution accounts for

1 Medical Research Center-UNAM-UABJO, Medicine Faculty, Oaxaca, Mexico

2 Forensic Science Doctorate. CADIT, Xalapa, Veracruz, Mexico

3 Faculty of Law, UABJO, Oaxaca, México.

4 Department of Biochemistry and Immunology, ITO-UNAM, Oaxaca, Mexico

Author Note

Correspondence concerning to this article should be addressed to Eduardo Perez-Campos Mayoral*, Medicine Faculty, UABJO, Oaxaca, Mexico D.F. Phone: +52 (951) 1257907

Acknowledgments

The authors thank Mark Handler for his thoughtful reviews and comments to earlier drafts and revisions to this paper. The authors thank Charlotte Grundy for her support throughout the work.



losses of about 1% of total sales.

It is widely documented that firms, on average, regardless of sector or country, lose about 3% of total annual sales due to employee and customer theft, and administrative errors (Ace Hardware Corporation, 2010). Many companies consider that dishonest employee conduct is virtually impossible to control and they regard inventory irregularities as administrative costs, thereby shifting those costs to customers. In fact, there has been considerable development in the field of Security and Criminology which deals with this dishonest behavior and provides businesses with specific tools to address the issue and reduce its impact (Cognizant 20-20 Insights, 2014).

Strategies to Halt Internal Theft

There are several preventative control measures that companies could adopt and implement as proactive steps.

Initiate controls through computerized inventory management systems to identify losses from receipt to point-of-sale, including warehouse storage and clearance of goods areas. These controls should maintain records for broken merchandise, losses and internal consumption, as well as regularly monitoring errors in price marking and coding. Additionally, an administrator level control point could help eliminate differences in stock and controls in cash management – i.e. implementation of vouchers and credit notes, limited authorization for purchase and sale adjustments, such as approval for sales cancellations or price markdowns. The systems could be set up to require a unique password so that users would be easily identified. Finally, a clear procedure for returns and exchanges should be implemented.

Management controls. Specifically, analyzing the critical points of the system. Efforts should be taken to strengthen controls and implement specific authorization or approval levels.

Pre-screening of prospective employ-

ees and a focus on maintaining high employee motivation. Managers and business leaders should be aware of the benefits of properly selecting and motivating staff, as it plays an important role in the improvement of business. The foundation is laid with the employee hiring process, though that is just the first step. There are various methods which could be implemented to evaluate attitudes and assess behavioral indicators. Businesses should manage and maintain a highly motivated staff through training or incentives. This would generate the commitment and productivity of employees which often results in a minimization of losses.

Companies could invest in a network of visible and hidden security cameras to reduce the occurrence of theft, along with displaying informative signs alerting customers and employees of ongoing surveillance. They would be useful for identification, and provide evidence for prosecution in any criminal charges and insurance claims should a theft occur.

By employing uniformed and undercover security personnel, along with “mystery shoppers”, the opportunity for employee theft is would be reduced. The mystery shoppers would be used to assess and measure the quality of customer service and can also help to identify corrupt employees through covert operations. The undercover operatives would be able to make scheduled purchases anonymously, to verify a correct records of sales by the cashier.

However, if preventative measures are ineffective and security has been compromised, it would be necessary to take proactive or confrontational action. That is to say, use methods of investigation to solve the problem, minimize losses, and, if possible, recover the stolen items and find the perpetrators.

All employees have some opportunity to engage in criminal behavior, and some follow through. According to international studies, 60% or more of workers have been involved in dishonest behavior (Zińczuk, et al., 2013). For example, in one company, 25% of



the employees regularly steal, 25% don't steal under any circumstances, and the remaining 50% will steal if given the opportunity to do so (Narasimham, 2005).

Elements contributing to employee theft have been extensively studied for decades and assessed in different ways. The consensus among researchers is best explained as to why employee theft occurs by the "Rob Triangle" theory. This triangle is composed of three basic elements: opportunity, necessity and behavioral rationalization (Hayes, 2008). The offender plays into two parts of the theory, that is, his or her "ability" or "desire" or perceived "need".

Need is defined as a feeling of lack, physical or psychological, which all human beings have, independent of their race, culture or religion. Even a person's economic status does not preclude them from falling prey to the drive of an unmet need, as is seen in cases from a wide variety of social and economic backgrounds. They are often individuals with strong purchasing power who have committed petty larceny, a misdemeanor (Krasnovsky & Robert, 1998). They justify their behavior to themselves and others as a means to meet their perceived needs.

Given the desire and ability to commit the crime, there only needs to be an opportunity of which the offender can take advantage. Neglect by the business provides a window of opportunity for a theft to occur without being discovered. Unfortunately, experience has shown us that a dishonest employee will often seek and find new ways of breaching security systems, either through the establishment of a criminal network or knowledge, and avoidance of various controls and security procedures. Invariably, human creativity surprises us, generating unthinkable solutions even the most qualified experts in the field, and surpassing the most sophisticated control systems.

It should be mentioned that differences between investigations in the public sector and the private sector are minimal. The private

sector has less unethical tendencies in than the public sector (Sardžoska, & Tang, 2009). The use of tools, techniques and processes is very similar. However, in the public sector, investigations are aimed towards prosecution of the person or persons responsible for the offense, whether it be theft against an individual or against society and the common good. Investigations conducted in the private sector are based on individual needs, particular situations, and the interests of each company or organization. A violation of a private sector's code of conduct or company rules may, at the same time, be a violation of the law, and thus, the initiation of an investigation by the authorities in the public sector would ensue. Theft is an example of what could begin with an internal investigation and lead to the involvement of public authorities for the appropriate enforcement of local and federal laws.

Polygraph Testing Use in Companies Overseas

In 1921 John Larson and Leonard Keeler developed the polygraph discipline (Goddard CH, 1976). The field has been plagued by internal debate as to which methods are the best. Some of those interested in developing this field established schools of polygraphy and promoted their own methodologies and techniques. This was closely linked to, and materially rooted in their own economic interests. Still, this field grew exponentially and soon polygraph testing could be found in use in every conceivable application.

Many things have changed since its invention, but its acceptance and utilization still varies around the globe. Nowadays, most examiners in the United States of America come from, or work for, the public sector, bringing a new spirit of a 'best practices' policy and a scientific methodology toward their work. Additionally, professional associations, such as the American Polygraph Association (APA, 2011), have tended to move towards a more balanced perspective in order to ensure that its members provide valid and reliable public services. All APA members are required to comply with the recommendations of APA and operate us-



ing standardized protocols. Polygraph testing is conducted in Canada and the United States of America with a strong emphasis towards public and private professional development. Yet there are still variances in the acceptance of this technology. In Canada, the results of examinations are inadmissible in the criminal justice system, as seen in Case 18856 (Glaney & Bradford, 2007). However, the results of a polygraph examination are admissible in a civil court or an employment tribunal. With regard to the law, the validity of such evidence is unclear. Civil courts across Canada have adopted a variety of views - ranging from the acceptance of the polygraph to partial or non-acceptance. In addition, confessions obtained after a failed Polygraph test are admissible in Canadian Criminal courts if the rules of admissibility have been respected, e.g. *Regina vs Oickle* (R v Oickle, 2000 SCC 38, 2 SCR 3). *R. vs Spencer* and *R. vs Singh*, are the leading cases in the Supreme Court of Canada on the Common Law Rule for Confessions (Dufraimon, 2008).

The United States created the Employee Polygraph Protection Act in 1988 which restricts the use of polygraph testing in the public sector, with some exceptions (Employee Polygraph Protection Act. [EPPA, 2012]), such as security service firms and companies that manufacture, distribute and dispense pharmaceutical products.

Further south in the Americas, polygraph testing is not fully regulated. The use of the polygraph in Latin America is continuously under attack by a critical academic and scientific community in numerous articles published on the subject. It is misused by corrupt officials and public servants in some countries and there is a widespread belief that the polygraph is illegal. The procedure has seen partial prohibition in Honduras (Meyer, 2015), Salvador (Bertelsmann, 2009), and recently in Mexico (Sabet, 2010), and an attempted prohibition in Ecuador (Ramírez & Balbina, 2015). In addition, other technologies have been introduced that are competing with the polygraph, such as EyeDetect® by Converus®, which offers an opportunity to increase the

outcome confidence of credibility assessment testing, thus improving lie detection.

In Mexico certain laws and institutional regulations govern and permit the use of polygraph testing with employees in Public Safety, Homeland Security, and Federal, State and Municipal departments of justice, but it is not regulated in private industry. In criminal trials in Mexico and other countries, the results of a polygraph test can be admitted as evidence (Comisión Nacional de los Derechos Humanos, CNDH 2004). However, even where the accused has given his or her consent, the following minimum standards must be observed during the first stage of questioning: a) the test must be done in presence of the judge who is obliged to receive all statements about criminal acts, whether they are offered as an expert opinion or not, b) as coordinator of the process, the Judge is obliged to ensure the correct presentation of the evidence, analyze the interrogation, and when appropriate, ask the expert to rephrase questions that are ambiguous, confusing, insidious, or are interpreted as misleading, the Judge also must have some form of training in Polygraph in order to understand the psychology behind the construction of the Comparison Questions, especially in the Probable Lie Test (PLT) (Handler & Nelson, 2009), and c) all questions and answers must be registered in the record of the indictment and prosecution with the objective of giving certainty to these statements and availability to the different parties for future reference and discussion (Suprema Corte de Justicia de la Nación, 2012).

Polygraph testing in court is a common practice in many countries throughout Asia (British Psychological Society, 2004). While small and medium sized companies in these countries make systematic use of the polygraph with the objective of identifying theft and fraud, their primary goal is theft prevention.

In various countries across Europe there is also ambiguity as to its use. For example, while polygraph application is frequently used in the judicial systems of Belgium and



Spain, it is not permissible as evidence in other countries (Meijer et al., 2008). In England, the police have begun using polygraph testing with suspected sex offenders. The test is carried out before the formal charges are issued and it is often useful for shortening the duration of a criminal investigation. Furthermore, polygraph tests are mandatory for sex offenders on parole in the East and West Midlands in England (Gannon et al., 2012). However, when it comes to its application in the business world, there is no common set of rules governing its use.

Polygraph in Private Sector

According to "The 2012 Marquet Report on Embezzlement" (Christopher, 2013) the number of major incidences of embezzlement increased 11% over 2011, and the average loss was approximately US \$1.4 million dollars. Employees who were working in finance, bookkeeping and accounting positions committed approximately 68% of embezzlements. Issuance of forged or unauthorized company checks was the most common embezzlement scheme in which employees were involved. The 2013 report detailed the continued rising trend, showing a further increase of 5% over 2012. Embezzlement by employees, as mentioned above, rose to 71%, 81 % of the cases being the number of individual perpetrators involved. In the last six years, California, Michigan, Pennsylvania, New York, Virginia and Texas have experienced the greatest number of major embezzlement cases in the USA.

In an effort to find possible solutions to this problem, most large chain stores in the US have developed their own Loss Prevention units. These units may employ an entire investigation team, including, as a minimum, a project manager, investigators, lawyers, support members and tasking executives with management responsibilities, making it a very expensive option. An alternative could be to outsource by hiring professional investigators who have experience and are skilled in interviewing and interrogation procedures such as the PEACE interview model, physical surveil-

lance, electronic surveillance, forensics, covert investigations and if company policies and local laws permit, the use of polygraph testing.

The American Polygraph Association (APA, 2011) established the use of validated techniques to eliminate the use of non-standardized techniques (Krapohl, 2006; APA, 2011). When polygraph testing is permissible, validated techniques indicate the necessity of performing some kind of preliminary investigation before administering it, according to the APA, which has established models and procedural policies, written proposals and scientifically validated methods of testing (Handler et al., 2009). These concepts arose to safeguard the lives of individuals and protect the professional integrity of workers in the private sector.

Every day more companies are protecting their assets by increasing their security through the use of the different methods, such as pre-employment integrity testing, and in the US with a Personnel Selection Inventory (Brown, et al., 1987). However, considering that employees who hold finance, bookkeeping and accounting positions are more likely to commit embezzlement, this group of workers could be evaluated by polygraph to promote transparency and preventative security within the company (Terris & Jones., 1982).

Any incident of theft or fraud, in addition to economic damage, causes problems and internal tensions that affect the entire company. Trust is lost, rumors abound, fear of being marked as a suspect arises, and until clarified, the situation is not resolved. The situation is extremely tense and uncomfortable for both the employer and employees. For this reason, the company should work to bring closure to the offense as quickly as possible.

The polygraph offers individuals an opportunity to prove their innocence, taking away pressure and suspicion, and allows the employer to potentially unmask the real culprit and possible accomplices. In this way, the conflict is resolved quickly and fairly for all, and the company can return to business



as usual. If a polygraph test is performed, false-positive results may be followed-up with alternative technologies like EyeDetect (Handler, 2016), or functional magnetic resonance imaging (fMRI) (Farah et al., 2014; Langleben et al., 2016) or to eliminate any doubt.

The polygraph can also serve to clarify cases of workplace harassment, including sexual harassment, resolve conflicts, determine responsibility, and further investigation of suspicious events. However, the Food & Allied Workers Union obo Kapesi mentions “A polygraph test on its own cannot be used to determine the guilt of an employee” for (le Roux, 2013).

The use of the polygraph in private organizations fulfills the same function as the anti-doping controls in sport. The real goal is not to catch anyone, but to deter those who seek to commit any abuses or irregularities,

such as distortion of expenses, personal trading commissions with suppliers, deliberate breach of rules and obligations, and theft.

For preventive purposes, it is extremely practical and profitable. An awareness of testing practices will also help ensure employees remain operating within the company's business conduct guidelines. Periodic evaluations are an excellent control mechanism for all employees, with direct or indirect access to company assets.

Maintaining a systematic approach of giving periodic polygraph examinations to all employees, conducted exclusively as a control measure of professional performance and honesty of workers, may help to circumnavigate any concerns related to invasion of privacy or discrimination. Taking action against crime is not simple, nor is it inexpensive, but ignoring it is much more expensive.



References

- Ace Hardware Corporation (2010) Inventory shrinkage. Facts, figures & strategies. Downloaded from www.ancellossprevention.com/.../Inventory%20Shrinkage%20-%20Reduce%20It1.pdf
- American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40, 196-305.
- Beedle, B. (2005). Employee theft how to fight a multi-billion dollar problem. *National Property Management Association (NPMA)*, 17, 22-24.
- Bertelsmann S. (2010). BTI 2010 — El Salvador Country Report. Gütersloh: Bertelsmann Stiftung, [Electronic version] 2/06/2016, available from http://www.bti-project.de/uploads/tx_itaowdownload/BTI_2010_El_Salvador.pdf.
- British Psychological Society. (2004). A Review of the Current Scientific Status and Fields of Application of Polygraphic Deception Detection. Report (26/05/04) from the BPS Working Party”, [Electronic version] 2/06/2016, available from <http://www.bps.org.uk>.
- Brow, T.S., Jones, J.W., Terris, W., & Steffy, B.D. (1987). The impact of pre-employment integrity testing on employee turnover and inventory shrinkage losses. *Journal of Business and Psychology*, 2, 136-149.
- Christopher, T.M. (2013). The 2013 Marquet report on embezzlement”, [Electronic version] 2/06/2016, available from <http://www.hartfordbusiness.com/assets/pdf/HB106961222.PDF>.
- Cognizant 20-20 Insights. Cognizant (NASDAQ: CTSI). (2014). Strategies to Mitigate Shrink in a Boundary-less Retail World. [Electronic version] 2/06/2016, available from <https://www.cognizant.com/InsightsWhitepapers/Strategies-to-Mitigate-Shrink-in-a-Boundary-less-Retail-World-codex897.pdf>.
- Comisión Nacional de los Derechos Humanos. CNDH. (2004). Recomendación general no. 6, sobre la aplicación del examen poligráfico. [Electronic version] 2/06/2016, available from http://www.stps.gob.mx/bp/micrositios/trasparencia_focalizada/Recomendacion6/REC_Gral_006.pdf.
- Dufraimon, L (2008). The Common Law Confessions Rule in the Charter Era: Current Law and Future Directions. *Supreme Court Law Review*, 40 S.C.L.R. (2d), 249-269.
- Employee Polygraph Protection Act. EPPA. (2012). [Electronic version] 2/06/2016, available <http://www.polygraph.org/section/resources/employee-polygraph-protection-act-eppa>.
- Farah, M.J., Hutchinson, J.B., Phelps, E.A., & Wagner, A.D. (2014). Functional MRI-based lie detection: scientific and societal challenges. *Nat Rev Neurosci*, 15, 123-31.
- Gannon, T., Wood, J., Pina, A., Vasquez, E., & Fraser I. (2012). The evaluation of the mandatory polygraph pilot. University of Kent. Ministry of Justice Research Series 14/12, [Electronic version] 2/06/2016, available https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217436/evaluation-of-mandatory-polygraph-pilot.pdf.
- Glancy, G.D., & Bradford, J.M. (2007). The admissibility of expert evidence in Canada. *J Am Acad Psychiatry Law*, 35, 350-6.



- Goddard, CH. (1976). Some reminiscences on early days of the lie detector. *Polygraph* 5, 252-265
- Handler, M., (2016, July). Understanding target base rates, successive hurdles & alternative technologies. Presented at the 39th Annual American Association of Police Polygraphists Annual Seminar, Uncasville, CT.
- Handler, M., Honts C., Krapohl, D., Nelson, R., & Griffin S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Criminal Psychology*, 24, 69-86.
- Handler, M. & Nelson, R., (2008). Utah approach to comparison question polygraph testing. *European Polygraph*, 2(2), 83-110.
- Hayes, R. (2008). CRISP report. Strategies to detect and prevent workplace dishonesty. ASIS International Foundation, Inc. [Electronic version] 2/06/2016, available <http://www.popcenter.org/library/crisp/dishonesty.pdf>.
- Krapohl, D.J. (2006). Validated polygraph techniques. *Polygraph*, 35,149-155.
- Krasnovsky, T., & Robert, L.C. (1998). Shoplifting: A review of the literature. *Aggression and Violent Behavior*, 3, 219-235.
- Langleben, D.D., Hakun, J.G., Seelig, D., Wang, A.I., Ruparel, K., & Bilker, W.B. (2016). Polygraphy and functional magnetic resonance imaging in lie detection: a controlled blind comparison using the concealed information test. *J Clin Psychiatry*. (In press 8 July 2016)
- Meijer, E. H., Koppen, V. P., Canter, D., & Zukauskienė, R. (2008). Lie detectors and the law: The use of the polygraph in Europe. In, Canter, D., Zukauskienė, R. (Eds.), *Psychology and law: Bridging the gap*, Ashgate Publishing, Editors.
- Meyer P, J. (2015). Honduras: Background and U.S. relations. [Electronic version] 2/06/2016, available <http://www.fas.org/sgp/crs/row/RL34027.pdf>.
- Narasimham, R. (2005). A Basket of good looking apples, but...The employees who perpetrate frauds, their footprints and fingerprints. In Yogeshwar, L., Prasanna, K., Yugandhar, B., Sandhya, Srikanth V., Narasimhan, R., Krishna, Chaitanya K., & Pradhan Sourav Kumar (Ed.), *Fraud in Banks*, The ICAI University Press.
- Le Roux, P.A.K. (2013). Polygraph testing and employment. The Labour Court reaffirms limited evidential value. *Contemporary labour law*, 22,61-63
- R. v. Oickle. (2000). 2 SCR 3, 2000 SCC 38. The Canadian Legal Information Institute (CanLII). [Electronic version] 25/07/2016, available <http://www.canlii.org/en/ca/scc/doc/2000/2000scc38/2000scc38.html>
- Ramírez, A., & Balbina R. (2015). Necesidad de reformar al Código de Procedimiento Civil, para que permita someter a la prueba del polígrafo, a los testigos en el sistema procesal civil Ecuatoriano. 2015. [Electronic version] 2/06/2016, available <http://dspace.unl.edu.ec/jspui/handle/123456789/8438>.
- Sabet, D. (2011). Police reform in Mexico: Advances and persistent obstacles". In, Olson, E.L., Selee



- A., & Shirk D. Shared Responsibility, Chapter 10. [Electronic version] 2/06/2016, available <https://www.wilsoncenter.org/publication/shared-responsibility>.
- Sardžoska, E., & Tang, T. (2009). Testing a model of behavioral intentions in the republic of Macedonia: Differences between the private and the public sectors. *Journal of Business Ethics*, 87, 495-517.
- Suprema Corte de Justicia de la Nación. México. (2012). Registro No. 2002 182. “Prueba en poligrafía en el juicio penal. Reglas mínimas que deberán observarse durante el desahogo de la fase de interrogatorio de quien aceptó someterse a ella”. Localización: [TA], 10a. Época; T.C.C.; S.J.F. y su Gaceta; Libro XIV, Noviembre de 2012; Tomo 3; p1925.
- Terris, W., & Jones, J. (1982). Psychological factors related to employees’ theft in the convenience store industry. *Psychological Reports*, 51, 1219-1238.
- Zińczuk, B., Cichorzewska, M., & Walczewski, M. (2013). The analysis of unethical behavior among employees in enterprises – A pilot study in the automotive Industry. Active citizenship by knowledge management & Innovation. Zadar, Croatia. [Electronic version] 2/06/2016, available <http://www.toknowpress.net/ISBN/978-961-6914-02-4/papers/ML13-249.pdf>



Bayesian Probabilities of Deception and Truth-telling for Single and Repeated Polygraph Examinations

Raymond Nelson[†] and Finley Turner[‡]

“He who refuses to do arithmetic is doomed to talk nonsense.”— John McCarthy¹

Abstract

Bayes' theorem is explained and applied to polygraph test results. Potential advantages of successive hurdles testing strategies are shown, including the use of Bayes' theorem to aggregate the results when examinees are subject to multiple tests regarding the same target issue. An advantage of a Bayesian approach to probabilistic results of polygraph examinations is that the interpretation of results from Bayesian inference can be both intuitive and practical. Mathematical examples are provided in 2x2 tables for single and repeated polygraph exams.

Keywords: Bayesian, Bayes' theorem, Bayesian inference, prior probabilities, base rates, incidence rates, a priori, posterior probability, a posteriori, diagnostic exams, screening exams, retesting, successive hurdles.

Probabilities of Deception and Truth-telling for Single and Repeated Polygraph Examinations: Bayesian Analysis aka Bayesian Updating

Introduction

Quiz 1: *Suppose we have a test for a disease. The test has a sensitivity level of .95, meaning that it can detect the presence of the disease in 95% of cases with the disease. The test also has a specificity rate of .95, meaning that it*

can determine the absence of the disease in 95% of the cases that do not have the disease. Suppose, for convenience, that the false positive and false negative error rates are both .05. We test a patient. The test result is positive and the level of statistical significance, expressed by the error rate, is .05. What is the probability that the patient has the disease?

Answer: *We don't know until we have more information about the reason for doing the test.*

¹ John McCarthy (1927-2011) was a computer scientist and cognitive scientist who helped develop the Stanford AI Laboratory and is credited with coining the term “artificial intelligence.”

[†] Raymond Nelson is a psychotherapist and field polygraph examiner who works in the area of sexual abuse and victimization. Mr. Nelson presently works as a researcher in the polygraph profession, and has published articles on many aspects of the polygraph test including polygraph techniques, test data analysis, psychology, physiology and other topics. Mr. Nelson lectures frequently on polygraph topics, is the director of an accredited polygraph training program, and is a research specialist for Lafayette Instrument Company (LIC). Mr. Nelson is a past President of the American Polygraph Association (APA) and currently serves as an elected member of the Board of Directors. The views and opinions expressed herein are those of the authors and not necessary those of the APA or LIC. Mr. Nelson can be reached at raymond.nelson@gmail.com.

[‡] Finley Turner is a polygraph examiner who conducts examinations for individuals, investigations and risk management screening in private practice in the United Kingdom. Mr. Turner's background is in Economics, and Psychometrics for which he holds a European Test User Certificate and a membership with the British Psychological Society Testing Centre. Mr. Turner can be reached at finturner@gmail.com.



Why?: Understanding the reason for doing the test will help us to understand the prior probability that the patient has the disease. The test result describes the test data, which is a proxy or model for the real universe that includes the patient who may or may not have the disease. Translating the test result into a real-world probability statement requires Bayesian analysis, also known as Bayesian updating. We need to state our knowledge of the prior probability, before the test, that the patient has the disease. The prior probability for a screening exam on a member of some population can be thought of as the base rate or incidence rate of the disease or problem among all members of the population. A diagnostic exam on a symptomatic patient will have a substantially greater prior probability of having the disease. Different prior probabilities will lead to different posterior probabilities associated with a positive or negative test result.

Quiz 2: Suppose we are conducting a screening test on a randomly selected member of the population and the incidence rate for the disease is .05 or 5%. Also suppose the screening test has a sensitivity rate of .95 and a specificity rate of .95. We conduct the test, and the test returns a positive result. Again the level of statistical significance is .05. What is the probability that the patient has the disease based on the screening test result?

Answer 2: 50%. Thus illustrating one of the reasons that screening test results may not be sufficient to make a diagnosis.

Why?: Because the prior probability or incidence rate is low. Using Bayesian analysis, we used the test result to update the prior probability of disease to a posterior probability of disease. Here is the math:

posterior probability of disease =

$$(\text{prior probability} * \text{test sensitivity rate}) /$$

$$((\text{prior probability} * \text{test sensitivity rate}) + ((1 - \text{prior probability}) * \text{false positive rate}))$$

Using the numbers, and flipping the result to the right side of the =, this means:

$$(.05 * .95) /$$

$$((.05 * .95) + ((1 - .05) * .05)) = .5$$

This is the reason that screening tests are often useful to suggest areas for further testing or investigation but are of themselves, insufficient to achieve a diagnosis.

Quiz 3: Now suppose we have completed a screening test on a patient who was selected for testing only because he is a member of the population (i.e., random selection), not because of a known problem. The incidence rate of the disease is still 5%. But we have completed a previous screening test that gave a posterior probability of 50% that the patient is positive for the disease. We then conduct a diagnostic test, which also, for convenience, has a sensitivity rate of .95 and also a specificity rate of .95, along with false positive and false negative error rates of .05. We use the posterior probability (.5) from the screening test as the prior probability for the diagnostic test. Suppose the test result is positive, and the level of statistical significance is .05. What is the probability that the patient has the disease?

Answer 3: 95%. Showing that satisfactory levels of test accuracy can sometimes be achieved through multiple testing strategies when the prior probability is low.

Why? The screening test increased the prior probability that the patient has the disease from 5% in the population



to 50% among those who produced positive results on the screening test. The greater prior probability gives us greater precision (i.e., a reduced margin of uncertainty and error) for the Bayesian posterior probability of disease when the two test results concur. Using the same formula as before with the prior probability of .5, here is the math:

$$(.5 * .95) /$$
$$((.5 * .95) + ((1 - .5) * .05)) = .95$$

The purpose of Bayesian analysis is to quantify the degree of belief or level of confidence we may have about some phenomena for which our knowledge is uncertain or imperfect. Of interest here is that statistical estimates such as test sensitivity, specificity and decision error rates do not always translate directly to conclusions about real-world effect sizes. However, Bayes' theorem and Bayesian analysis (Berger, 1985; Gelman, 2014; Winkler, 1972) can be used to calculate posterior probabilities that are conditional on prior probabilities (Cohen, 1994) and which are intended to describe real world effects. Bayesian inference has important practical usefulness when attempting to understand real-world probability problems such as when attempting to measure or quantify amorphous phenomena, such as deception and truth-telling, that cannot be subject to deterministic observation or direct physical measurement.

What is Bayesian Analysis, and What Does It Have to Do with Polygraph?

The theory of the polygraph holds that greater changes in physiological activity can be observed in response to different types of test stimuli as a function of deception or truth-telling

regarding the relevant target stimuli (American Polygraph Association, 2011; Honts & Peterson, 1997; Kircher & Raskin, 1988; Nelson, 2015a, 2015b; National Research Council, 2003; Office of Technology Assessment, 1983; Senter, Weatherman, Krapohl & Horvath, 2010). Although deception and truth-telling cannot themselves be subject to direct physical measurement or deterministic observation, certain changes in physiological activity have been shown to be correlated with deception and truth-telling at statistically significant levels (Bell, Raskin, Honts & Kircher, 1999; Harris, Horner & McQuarrie, 2000; Kircher, Krisjansson, Gardner & Webb, 2005; Kircher & Raskin, 1988; Podlesny & Truslow, 1993; Raskin, Kircher, Honts & Horowitz, 1988). Changes in physiological activity can be numerically coded and combined in a structured algorithm to calculate a numerical score that can be compared to empirical or theoretical distributions to calculate a statistical classifier. Bayesian analysis can combine information from a statistical classifier with a prior probability and the result can be interpreted as a posterior probability of deception or truth-telling.

Bayesian analysis refers to a particular method or style of probabilistic thinking, named for Thomas Bayes (1701-1761). Bayes was a Presbyterian minister who was interested in probability theory. He was important during his lifetime, and perhaps even more important in the centuries since then. He was elected to membership in the Royal Society² in 1742. *Bayes' theorem*, as it is known today, was not published during Bayes' lifetime. It was contained in an essay that was discovered

2 The Royal Society was chartered by King Charles II in 1660, and serves today as the national academy for the sciences to the United Kingdom. The Royal Society published the world's first scientific journal, *Philosophical Transactions*, in 1665 and established the fundamental principle and practice of peer review.

3 Richard Price, mathematician, moral philosopher, and protestant minister, was also a friend to the Founding Fathers of the United States, and a member of the Royal Society to which American Statesman Benjamin Franklin (1706-1790) was elected to membership in 1756.



after his death. A friend named Richard Price³ was executor to Bayes' intellectual work following his death and recognized the importance of the Bayes' idea. It was Price who arranged for the reading and publication of Bayes' essay by the Royal Society (Bayes & Price, 1763; Laplace, 1812)⁴. Publication of a new text book, *Theory of Probability* by Harold Jeffreys⁵ (1939) reintroduced Bayesian analysis to 20th century readers.

In contrast to the Bayesian paradigm is the *frequentist* statistical approach - the other major style of statistical or probabilistic thinking that emerged as central to the 20th century practice of null-hypothesis significance testing - based largely on the work of Ronald Fisher⁶ (1925; 1935), Jerzy Neyman⁷ and Egon Pearson⁸ (Neyman & Pearson, 1938; 1933). Under the paradigm of frequentist inference a hypothesis is tested with no explicit assumption of a prior probability, as if we have no prior knowledge or assumption about the probabilities associated with the different possible outcomes. Instead, the frequentist approach is to attempt to distill a world full of messy and incompletely controlled data to a simple go/no-go dichotomous classification regarding the level statistical significance of

a hypothesis or conclusion. This is accomplished by declaring a probability decision threshold in advance (sometimes referred to as *alpha* or the significance level), representing a tolerance for uncertainty and error. A null hypothesis (or hypothesis of no effect) is never actually proven, but can be falsified and rejected. In practice this has sometimes resulted in misunderstanding and misrepresentation of the practical meaning of the categorical and probabilistic results from null-hypothesis significance testing. For example, the measure of statistical significance, the p-value, does not allow one to find the complimentary probability of a correct conclusion.

Jeffreys (1980) criticized the frequentist practice of null-hypothesis significance testing as basing a conclusion on an attempt to disprove that something has not occurred. Most controversy surrounding null-hypothesis significance testing hinges on misunderstanding and misapplication of the procedure and around the practice of point null-hypothesis testing⁹ and dichotomous accept/reject procedure (Berger & Sellke, 1987). Another concern is that frequentist p-values can be smaller than Bayesian probability estimates for some prior probability distributions (Berger & Delampady, 1987). Johnson (2013) showed that

4 Simon Pierre Laplace (1749-1827), one of the greatest scientists in history, also played an important role in the development of probability theory and what is today known as *Bayes' theorem*, having published *Théorie analytique des probabilités* (1812). Laplace's work in calculating the locations of celestial objects led to greater understanding of measurement error and also foundational in the development of the principles of frequentist inference, including p-values, probability distributions and the central limit theorem.

5 Harold Jeffreys was a British geophysicist, astronomer, mathematician and statistician who was interested in Bayes' theorem. He was elected to membership in the Royal Society in 1925. As a geophysicist, Jeffreys suggested that the earth's planetary core was liquid, and was opposed to the theory of continental drift because of the lack of a sufficient motive force (the larger theory of plate tectonics has now subsumed continental drift, and supplies an accepted theory of the movement of the earth's surface).

6 Ronald Fisher was famous as a biologist who studied genetics and evolution, as an agricultural scientist who studied crop variation, and as a statistician who anchored our understanding of frequentist statistical methods. Fisher was elected to membership in the Royal Society in 1929.

7 Jerzy Neyman was a Polish statistician who moved to California in the late 1930s. Neyman developed the department of statistics at the University of California at Berkeley. Neyman introduced the concepts of confidence intervals and stratified sampling. Together with Egon Pearson, Neyman developed the concept and practice null-hypothesis significance testing (Neyman & Pearson, 1928).

8 Egon Pearson was a British statistician (whose father was Karl Pearson to whom we owe our understanding of the correlation coefficient). Pearson was elected to the Royal Society in 1966. Pearson, along with Neyman, strongly influenced the practice of null-hypothesis significance testing as we know it today.

9 With a sample of sufficient size two points are always different at a statistically significant level.



frequentist statistical results can be mathematically reconciled with Bayesian results, and argued that non-reproducibility problems related to testing at unjustifiably high levels of significance could be remediated by testing at $p=.005$ or $p=.001$ instead of $.05$ and $.01$. Most enduring criticism of null-hypothesis significance testing has centered largely on misuse and misinterpretation of results, and what has been called the “weak form” (Cohen, 1994) in which one attempts to confirm a theory by rejecting a null hypothesis.

In contrast to this, the “strong form” of null-hypothesis significance testing has not been subject to the same concern. The strong form of null-hypothesis significance testing involves taking a predication or hypothesis and then designing a test or experiment to challenge and discard it if possible, in a manner consistent with Karl Popper¹⁰ (1959) who asserted that falsifiability is a central requirement of scientific statements. Casella and Berger (1987) showed that p-values can be mathematically reconciled with Bayesian results, and concluded that one-sided p-values can be regarded as equivalent for some classes of prior distributions and within the range or at the lower bound of Bayesian measurement for others. Hwang *et al.* (1992) showed that Bayesian analysis did not outperform p-values for one-sided tests, leading Casella and Wells (1993) to conclude that p-values remain a viable assessment of the plausibility of a null hypothesis. Meehl¹¹ (1967, 1997), who credited David Lykken¹² with stimulating his thinking in this area, asserted the math-

ematical and statistical calculations of the of null-hypothesis significance test to be sound while the application and practice is sometimes problematic.

Bayesian statistical methods, because they are intended to test the hypothesis of interest directly as opposed to the null hypothesis, avoid much of the confusion and troublesome impulses surrounding null hypothesis significance testing. Bayesian probability models begin with a declaration of our prior knowledge or assumption of the probabilities associated with different possible outcomes. This prior probability can be thought of as a probability distribution. In the case of outcomes the prior probability distribution can be thought of as complimentary values that describe the proportions of deceptive and truthful persons. Prior probabilities are combined, using Bayes’ theorem, with the observed data and evidence from a test or experiment. The end result is a posterior probability that is intended to be more precise, or less incorrect, than the prior probability.

When the prior probability is estimated from the empirical observation of available data, and then conditioned on probability result or likelihood function from a test or experiment, the method can be referred to as *Empirical-Bayes*. In contrast, a *Subjective-Bayesian* approach would have us use any source of information, perhaps including prior information from unreproducible sources. [See Casella, (1985; 1992) and (Carlin & Louis), 2000 for more information on Bayesian anal-

10 Karl Popper was an Austrian-British philosopher of science who immigrated to New Zealand in 1937 and later to the United Kingdom. Popper was elected to the Royal Society in 1976. In his early years Popper worked variously in construction, cabinetmaking, and also a psychoanalyst for children, before completing a dissertation in what might be thought of today as cognitive psychology. After noting the contrast between the falsifiable assertions of Einstein’s theory of relativity (and Einstein’s statement that he would reject his own theory if it failed certain tests) and the untestable and unfalsifiable notions of Freudian psychoanalytic psychology - in addition to the seemingly unquestionable political dogma that was emerging in in some regions during the early 20th century - Popper contributed enormously to our notion of falsifiability and the need to question evidence and assumptions as a central requirement for science.

11 Paul Meehl (1920-2003) was a professor of psychology at the University of Minnesota. His work is also important in the philosophy of science and the falsifiability of scientific statements. He is one of the most cited researchers of the 20th century. Meehl theorized in 1962, while serving as president of the American Psychological Association that schizophrenia has a genetic link. He also helped in the development of the MMPI. In 1954, Meehl’s book showed the superiority of formalized algorithmic analytic protocols over unstructured expert clinical judgment, and permanently altered the discussion and direction of testing and diagnostic work that can involve both subjective clinical evaluation and test data.

12 David Lykken (1928-2006) was a professor of Psychology and Psychiatry at the University of Minnesota, known for his work in both twin studies and lie detection. He published critical views of the probable lie comparison question and its use in the polygraph test.



ysis.] Regardless of whether objective or subjective sources of information are used for the estimation of a prior probability distribution, the essence of Bayesian analysis is that prior knowledge, information, or belief can be quantified probabilistically and then mathematically updated with new data or information from a test or experiment. All probabilistic information is imprecise or imperfect. The resulting posterior distribution is intended only to be less incorrect and therefore a more precise estimate of the quantity of interest.

Bayesian analysis offers important practical value to the polygraph test because the results from Bayesian analysis of polygraph data can be interpreted intuitively as a *probability of deception*. Bayes' theorem offers an additional advantage to the polygraph context when conducting repeated examinations on the same examinee because results from a previous polygraph test can be used, like any useful information source, as a basis with which to estimate the prior *probability of deceptive or truth-telling*.

It has been over two and one-half centuries since Thomas Bayes' essay was read and published. Since that time it has become central to many discussions in statistical classification, inference and prediction. Practical interest in Bayes' theorem will likely remain strong well into the 21st century, including in the polygraph and others scientific contexts that involve inference, prediction and classification.

Discussion

What Is Bayes' Theorem, and What Does Posterior Probability Have to Do with Polygraph?

Bayes' theorem is useful in the polygraph context because the posterior probability estimate can be reasonably interpreted as a probability of deception or truth-telling. Bayes' theorem, expressed mathematically, is shown in Figure 1. Bayes' theorem is used to describe the probability of an event or condition based on other associated conditions or information. A theorem is a statement or proposition to describe how something (some small part of the universe) works. Theorems are often not self-evident and are subject to mathematical

and logical proof before they are accepted as correct. A posterior probability (*a posteriori*) is the resulting probability estimate when an assumed prior probability (*a priori*) is algebraically updated using Bayes' theorem and the results from a test or experiment.

Figure 1. Bayes' theorem.

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

$P(A|B)$ is the posterior probability of event A given condition B. $P(A)$ is the prior probability of event A. $P(B|A)$ is the probability of event B given condition A, and $P(B)$ is the prior probability of event B. Posterior probability can also be referred to as *conditional probability*, because the resulting probability estimate is conditional on some other probability event. Reading Bayes' formula we would say this: the posterior probability of event A given condition B is equal to the prior probability of event A multiplied by the prior probability of event B given condition A, divided by the prior probability of event B.

Bayes recognized that we do not begin any experiment, test or data analysis with zero knowledge or no assumptions about what the data might mean. Instead we enter every probability estimation task with some prior knowledge or assumption. Instead of denying our prior knowledge with the pretense that we can somehow know nothing, a Bayesian view holds that the purpose of a test or experiment was to improve our prior knowledge through the analysis of data, and that the improved or updated knowledge will remain fundamentally probabilistic.

Whereas the practice of null-hypothesis significance testing is an attempt to answer a question about the probability of obtaining the observed data if the null-hypothesis is correct, Bayesian probabilities provide an answer to the more practical question about the probability that the null-hypothesis is correct after obtaining the observed data. To accomplish this Bayes' theorem can be restated or re-written to describe the probability that a hypothesis or conclusion is correct by simply substituting H (hypothesis) for A and D (data)



for B. Figure 2 shows Bayes' theorem in terms of a hypothesis or conclusion

Figure 2. Bayes' theorem stated in terms of a hypothesis.

$$\text{Posterior}(H|D) = \frac{\text{Prior}(H) * P(D|H)}{\text{Prior}(D)}$$

Posterior($H|D$) is the posterior probability that a hypothesis is correct after obtaining the observed data. $P(H)$ is the prior probability that the hypothesis is correct. $P(D|H)$ is the probability of obtaining the observed data if the hypothesis or conclusion is correct. This will be the test sensitivity level for deceptive classifications or the test specificity level for truthful classifications. $P(D)$ is the prior probability of obtaining the observed data. This will be either the sum of the test sensitivity and the FP error rates, or the sum of the test specificity and the FN error rates. Reading the formula we would say this: *the posterior probability that the hypothesis or conclusion is correct given the observed data is equal to the product of prior probability that the hypothesis is correct and the probability of obtaining the observed data if the hypothesis is correct, divided by the prior probability of obtaining the observed data.*

Bayes' theorem can be used to algebraically transform a frequentist p-value into a posterior probability that a hypothesis or conclusion is correct (Cohen, 1994). This makes Bayes' theorem potentially useful in the polygraph context, and many other contexts, because it can be used to calculate the probability that a positive or negative test result is correct or incorrect. In the polygraph context, results of Bayesian analysis can be interpreted as a probability of deception or truth-telling.

To implement Bayes' theorem it will be necessary to know some of the test characteristics, including test sensitivity, test specificity, FP and FN error rates. It will also be necessary to clearly state our assumptions about the prior probability. Because the assumed prior probability impacts the resulting poste-

rior probability, selecting or defining the prior probability is an important problem in Bayesian analysis. Much has been written about the topic of prior probabilities, including descriptions of different bases for the construction or selection of priors.

One possible solution, when there is little information about the prior probability distribution, is to set the probabilities as uniform for all possibilities. When there are two possible outcomes the prior probability will be 50%. This is sometimes referred to as an *objective* prior, or *Objective-Bayes*, because it attempts to remove subjectivity by regarding all possible outcomes as equally possible. Use of a uniform prior probability distribution has also been referred to as a *non-informative* prior or *weak prior* because the prior has little impact on resulting posterior probabilities that are primarily determined by the test statistic under this condition. (For more information on prior probability distributions, the reader is directed to more exhaustive sources including: Kass & Wasserman, 1996; Box & Tiao, 1993; Bernardo & Smith, 1994; and Pericchi & Walley, 1991.)

Another possible solution, instead of either guessing or attempting to assume we know nothing, is to use data to derive an empirical estimate the prior probability distribution. This can be as simple as selecting a prior probability for a criminal investigation based on the number of possible suspects. For example: investigation of a theft involving only two possible suspects would also invoke a prior probability of .5 (50%). Of course, in this example the empirically derived prior is equal to the uniform prior.

Applied to the polygraph context, a goal of scientific testing and Bayesian inference will be to improve the proportion of correct classifications of deception and truth-telling compared to the prior probability. Although a test has merely to exceed the 50% chance level to be considered effective when the prior probability is assumed to be uniform for two possible outcomes, optimization of a test for practical usefulness under other prior proba-



bility distributions will depend on the test sensitivity, specificity and error rates that more than minimally exceed chance levels. Table 1

shows the test sensitivity, specificity, FP and FN rates reported by APA 2011.

Table 1. Polygraph accuracy (APA 2011, shown in Table 2 on page 237).

Test sensitivity or true positive rate (TP)	.812
Test specificity or true negative rate (TN)	.717
False positive rate (FP)	.144
False negative rate (FN)	.083

Following the example in Cohen (1994), we re-write Bayes' theorem while making three substitutions: 1) $P(H)$ is replaced by the prior probability that a case is actually positive, 2) $P(D|H)$ is replaced with the test sensitivity level, and 3) $P(D)$, the prior probability of obtaining positive test data, is replaced with the sum of two values: the product of the FP error rate and the prior probability that the case is actually negative, and the product of the test sensitivity rate and the prior proba-

bility that the case is actually positive. Appendix A shows the software code for an R 3.2.1 function (R Core Team, 2015) to compute a posterior probability. All computations were completed using Rstudio 0.99.467 (RStudio Team, 2015). Figure 3 shows Bayes' theorem written in terms of a posterior probability of deception given a positive test result using a uniform prior probability of 50% for deception and truth-telling.

Figure 3. Bayes' theorem re-written in terms of a posterior probability of deception.

$$Posterior(D|pos) = \frac{(.5 * .812)}{(.5 * .812) + (.5 * .144)} = \frac{.406}{.406 + .072} = \frac{.406}{.478} = .849$$

Also following the example provided by Cohen (1994), A probability of truth is shown in Figure 4, using slightly different substitutions: 1) $P(H)$ is replaced with the prior probability assumption that an examinee is truthful, 2) the test specificity level is substituted for $P(D|H)$, or the probability of truthful test data if the examinee is actually truthful, and

3) $P(D)$, the prior probability of obtaining truthful test data, is calculated as the of sum of two products: the FN error rate multiplied by the prior probability of deception, and the test specificity rate multiplied by the prior probability of truth-telling. Figures 4 is also written with a uniform prior probability.

Figure 4. Bayes' theorem re-written in terms of a posterior probability of truth-telling given a negative test result.

$$Posterior(T|neg) = \frac{(.5 * .717)}{(.5 * .717) + (.5 * .083)} = \frac{.358}{.358 + .042} = \frac{.358}{.400} = .895$$

It is reasonable to interpret the results of a Bayesian experiment as a probability of deception or truth-telling because the algebraic formula begin with and includes our existing information about the prior probability of deception or truth-telling. The prior probability is conditioned on the test statistic. Figures 3 and 4 show the posterior probability of deception to be $\sim .85$, while the posterior probability of truth-telling is $\sim .90$ when we use a prior probability of 50%. Bayes' theorem is simply a function to update a prior probability with the results from a probabilistic test or experiment.

Where Do Prior Probabilities Come From?

Prior probabilities can be derived from a number of information sources, including surveys, historical records, program evaluation reports, other scientific studies. Prior probabilities describe our knowledge about the likelihood of different possible outcomes prior to a test or experiment. Prior probabilities also describe the expected rate of correct and incorrect classifications. An *Empirical-Bayes* approach can make use of *base-rates* or *incidence rates* as a prior probability distribution. However, it is the prior probability estimate that is of importance here. The base-rate is simply used as a reasonable (i.e., evidence-based and therefore less subjective or arbitrary) estimate for the prior probability.

A *Subjective-Bayes* approach to prior probabilities would permit the use of expert intuition, random chance, or any other heuristic solution as a prior probability estimate. Use of subjective prior information permits the potential use of Bayes' theorem to address different types of problems, but is not without hazard and controversy. For example: in the case of a criminal investigation for which the investigator feels 99.9% sure a particular suspect is guilty the test result will be loaded for a positive outcome regardless if the suspect is innocent. For this reason, use of Empirical-Bayes and objective priors may be preferable to Subjective-Bayes.

What Happens When the Prior Probability Is Different?

When the prior probability estimate is low it can be expected that the proportion of false positive (FP) results is greater than the proportion of true positive (TP) results. To simplify discussion and illustration of Bayesian posterior probability calculations, data can be shown using a 2x2 table using the expected frequencies for 1000 cases. Table 2 shows the results using the test sensitivity (true positive, TP), specificity (true negative, TN), false positive (FP), and false negative (FN) rates shown in Table 1 along with a prior probability estimate of 10%.

Table 2. Table summary of the Bayesian posterior probabilities for N=1000 with prior probability = 10%.

	Guilty(deceptive)	Innocent(truthful)	Totals* (expected value)	Correct classifications**	Posterior Probabilities
Positive(deceptive)	.812*.1*1000=81(TP)	.144*.9*1000=130(FP)	211	81	.384
Negative(truthful)	.083*.1*1000=9(FN)	.717*.9*1000=645(TN)	654	645	.986
-	-	-	865	726	.839
* Column totals do not sum to 1000 because some polygraph examinations results are not statistically significant for deception or truth-telling and are therefore inconclusive and unclassifiable.					
** Correct classifications are rounded down to the floor integer. Incorrect classifications are rounded up to the ceiling integer.					



When the prior probability of deception is 10% the posterior probability of truth-telling for negative test results is approaching 99%. In the example shown in Table 2, 13.5% or 135 of the 1000 cases were inconclusive. Of the 654 cases that produced negative results only 9 were errors. However, 130 (61.6%) of the 211 positive test result were FP errors under the 10% prior probability condition. A comparatively low prior probability of 10% can also be referred to as a *strong prior* because of the dominating effect this prior probability has when combined with the other information. This has also been referred to as an *informative prior* for the same reason.

What Does All of This Have to Do With Re-Testing or Repeating a Polygraph?

Repeated testing practices can include any field practice in which more than one polygraph test is conducted on the same examinee. Reasons for the repetition of a polygraph examination may vary, but will generally involve the objective of ensuring the accuracy or usefulness of the test result. The simplest reason for repeating a polygraph examination is when the probabilistic test result does not meet required thresholds for statistical significance and so does not support an evidence-based opinion or conclusion.

Repeated polygraph tests can occur in the criminal investigation context when a subsequent polygraph test is conducted regarding an alleged crime or incident for which one or more polygraph tests has already been conducted on the same examinee, for example: if new information has become available. Repeated polygraph tests might also occur in circumstances in which attorney makes a referral for a private polygraph test prior to advising a client to present oneself for a polygraph test conducted by law enforcement

investigators. Some examinees who undergo polygraph screening may be subject to repeat testing or re-examination in order to ensure that test results, intended for risk management, describe the examinee's truthfulness, behavior, or integrity during the recent period of time. A potential benefit of repeat testing is the opportunity to combine the results of two polygraph examinations using the laws of probability theory¹³.

Repeated testing practices can be observed in the successful hurdles approach to polygraph screening (Krapohl & Stern, 2003). In the successive hurdles context negative results are considered conclusive while positive results from an initial screening test will be subject to additional testing before the results can be regarded as a basis of information for decision or action. In the ideal field practice scenario all subsequent testing would be conducted with a different type of test, using different instrumentation and analytic methods. The reason for this is that if there is a systematic (i.e., non-random) cause for testing error those same conditions could be present at both the initial and subsequent examination. Use of the same test method (e.g., polygraph and polygraph) will mean that both stages of testing may have similar vulnerabilities and similar potential for error. Use of different testing and analysis methods will mean that the two stages of the successive hurdles testing model are more likely to present different vulnerabilities to potential testing error.

Results from multiple examinations in successive hurdles testing strategies, in which two polygraph exams are conducted on a single individual, can be combined using Bayes' theorem. As previously indicated, prior probabilities can be derived from several different types of information, including surveys, historical records, program evaluation reports,

13 Paired testing practices – in which polygraph examinations are administered to two opposing witnesses regarding a single allegation or incident – can make use of the probability multiplication rule if the examinations are independent examiners who are blind to both the case status and the other test result. In the paired testing context it is the issue and not the examinee that is tested repeatedly.



other scientific studies, professional intuition, or even random chance. Results from previous testing can also serve as a basis for a prior probability estimate to calculate an improved probability estimate for the result of a subsequent test. This means that results from initial polygraph test might be used to improve the precision and effectiveness of a subsequent polygraph test.

What Happens to Test Accuracy if We Conduct Two Examinations on Everyone?

Table 3 shows a 2x2 summary table for the first of two examinations using the values

in Table 1 with $N=1000$ cases consisting of 500 guilty and 500 innocent cases^{14, 15}. A total of 478 cases are expected to produce positive results, for which 406 (84.9%) can be expected to be correct. A total of 400 cases are expected to produce negative results, for which 358 (89.5%) can be expected to be correct. As shown in Table 3, 764 (87.0%) of the 878 classifications can be expected to be correct. This is similar to the .869 estimate reported by APA (2011), as are the 122 inconclusive results that make up the 12.2% inconclusive rate.

Table 3. Table summary of Bayesian posterior probabilities for $N=1000$ with prior probability = 50%.

	Guilty(deceptive)	Innocent(truthful)	Totals* (expected value)	Correct classifications**	Posterior Probabilities
Positive(deceptive)	.812*.5*1000=406(TP)	.144*.5*1000=72(FP)	478	406	.849
Negative(truthful)	.083*.5*1000=42(FN)	.717*.5*1000=358(TN)	400	358	.895
-	-	-	878	764	.870
* Column totals do not sum to 1000 because some polygraph examinations results are not statistically significant for deception or truth-telling and are therefore inconclusive and unclassifiable.					
** Correct classifications are rounded down to the floor integer. Incorrect classifications are rounded up to the ceiling integer.					

Table 4 shows the Bayesian posterior probabilities for subsequent polygraphs conducted using the posterior probabilities from Table 3 as the prior probabilities for deception and truth-telling when the results of the two examinations concur. The prior probability of deception was increased from 50% to 84.9% when re-testing cases that had previously produced a deceptive test result. Similarly, a

truthful test result at an initial exam has similarly established a greater prior probability that an examinee is actually truthful when re-testing a case that produced a negative result at a previous polygraph test. Results shown in Table 4 were calculated with the assumption that the results of the two exams are independent given the criterion state¹⁶.

14 Guilt and innocence are legal terms that are used here only to differentiate the case status.

15 Though it sometimes occurs that the prior probability is an estimate of an unknown parameter, in this example the prior probability is treated as a known prior.

16 Independence among repeated polygraph test results has not been fully investigated. Repeated polygraph exams may or may not be completely independent in the manner of Bernoulli trials (i.e., coin tossing). If there is any source of shared variance between first and second examination (e.g., characteristics of the examinee in addition to the criterion state) then these calculations might represent the upper and lower bounds of the joint probability distribution after the second exam.



As expected, overall decision accuracy was observed to be greater when the results of the initial and subsequent examination are in agreement. Of the 340 positive classifications at the second exam, 329 (96.8%) can be expected to be correct, while 256 (98.5%) of the 260 negative classifications are expected to be correct. Overall decision accuracy is reduced when using concordant results from two examinations to classify the test results. In the example shown in Table 4 the number of FP errors was reduced to 11 (3.2%), and the

number of FN errors was reduced to 4 (1.5%) when a positive or negative classification depends on two concurring examination results. Table 4 also indicates that a large proportion of cases can be expected to remain unresolved by the requirement that the two examinations results concur. Unresolved cases will be a combination of the 122 cases that produced inconclusive results at the initial examination plus those cases that produced inconclusive results at the subsequent test, plus all cases for which the two test results did not concur.

Table 4. Table summary of Bayesian posterior probabilities for subsequent examination results that concur with the initial test result using the posterior results from an initial examination (shown in Table 3) as the prior probability.

	Guilty(deceptive)	Innocent(truthful)	Totals (expected value)	Correct classifications*	Posterior Probabilities
Positive	$812 \cdot .849 \cdot 478 = 329(\text{TP})$	$.144 \cdot .151 \cdot 478 = 11(\text{FP})$	340	329	.968
Negative	$.083 \cdot .105 \cdot 400 = 4(\text{FN})$	$.717 \cdot .895 \cdot 400 = 256(\text{TN})$	260	256	.985
-	-	-	600	585	.975

* Correct classifications are rounded down to the floor integer. Incorrect classifications are rounded up to the ceiling integer.

Table 5 shows the posterior probabilities when *all* results from the second examination are included in the analysis regardless of whether or not they concur with the results of the initial examination. Of the 879 conclu-

sive results, 672 (86.7%) can be expected to be correct with 121 (12.1%) cases inconclusive at the second examination. These results are within rounding error of the posterior probability for a single exam (shown in Table 3).

Table 5. Table summary of Bayesian posterior probabilities for subsequent examinations, using the results from Table 3 as the prior probability estimate, including test results that do not concur with the initial result.

	Guilty(deceptive)	Innocent(truthful)	Totals (expected value)	Correct classifications*	Posterior Probabilities
Positive	$812 \cdot .849 \cdot 478 = 329(\text{TP})$	$.144 \cdot .151 \cdot 478 = 11(\text{FP})$	340	329	.968
Negative	$.083 \cdot .105 \cdot 400 = 4(\text{FN})$	$.717 \cdot .895 \cdot 400 = 256(\text{TN})$	260	256	.985
Positive	$.812 \cdot .105 \cdot 400 = 34(\text{TP})$	$.144 \cdot .895 \cdot 400 = 52(\text{FP})$	86	34	.395
Negative	$.083 \cdot .849 \cdot 478 = 34(\text{FN})$	$.717 \cdot .151 \cdot 478 = 52(\text{TN})$	86	52	.605
Positive	$.812 \cdot .105 \cdot 500 = 42(\text{TP})$	$.144 \cdot .139 \cdot 500 = 11(\text{FP})$	53	42	.792
Negative	$.083 \cdot .105 \cdot 500 = 5(\text{FN})$	$.717 \cdot .139 \cdot 500 = 49(\text{TN})$	54	49	.907
-	-	-	879	672	.867

* Correct classifications are rounded down to the floor integer. Incorrect classifications are rounded up to the ceiling integer.



There were 43 FN errors, which was 8.6% of the 500 guilty cases, along with 74 FP errors, 14.8% of the 500 innocent cases, when using all results from the second examination. Excluding the 122 inconclusive cases that produced inconclusive results at the first examination resulted in 772 conclusive results for which 671 (86.9%) can be expected to be correct. No increase in decision accuracy can be expected when conducting two examinations unless the test results concur for the two examinations.

In practice, working at the level of the individual examinee, a requirement for two concordant exam results must be made in advance of the test administration and test data analysis. Otherwise the results of the first examination are at risk for influencing decisions about how to manage and prioritize test accuracy, and the effect can be a subsequent manipulation of test accuracy outcomes. When the deceptive and truthful classifications are made only when the results of the two examinations concur the result can be a substantial increase in decision accuracy. Another obvious practical issue is that achievement of high accuracy by requiring two concurring examination results will effectively double the costs of testing in terms of personnel, time, financial and physical resources.

What Happens When Retesting Inconclusive Results from the Initial Exam?

Results from re-testing the 122 cases that are expected to produce inconclusive results at the initial examination are included in Table 5. Of the 54 expected truthful classifications at the second examination following a previously inconclusive result 49 (90.7%) are expected to be correct with 5 (9.3%) FN errors. For the 53 expected deceptive classifications following a previously inconclusive result 42 (79.2%) can be expected to be correct with 11 (20.8%) FP errors. Decisions, not including inconclusive results, for guilty cases resulted in 42 (89.4%) correct of 47 classifications, and 5 (10.6%) FN errors. For innocent cases there were 49 (81.7%) correct of 60 classifications and 11 (18.3%) FP errors. Of the 53 positive classifications, 42 (79.2%) were correct, while 49 (90.7%) of 54 negative classifications were correct when retesting previously inconclusive cases. Fifteen cases, 12.3% of the previ-

ously inconclusive cases remained inconclusive after the second test. This was 1.5% of the original 1000 cases. Overall, 91 (85.0%) of 107 conclusive results were correct when inconclusive cases were re-tested. For practical purposes, test accuracy at the second examination can be expected to be the same as for the initial examination if all test results are to be considered.

What Happens if We Do the Second Test Only After a Positive Result at the First Test?

A possible alternative to the cost of conducting two polygraph tests on every examinee will be to conduct a subsequent examination only following a positive result at the initial examination. This type of testing strategy may be useful in polygraph screening programs in which the results of polygraph screening tests are indicative of a possible problem but are not of themselves sufficient as a sole basis for action. Negative test results in this scenario would simply be accepted at the known level of precision for a single examination. FN errors could be constrained through the selection of probability cut-scores and decision rules that constrain errors to within an established tolerance.

Table 6 shows the posterior probability for deceptive test results when testing only those persons that do not pass the initial examination (shown in Table 3). As shown earlier, 340 positive test results can be expected of the 478 cases that produced positive results at the initial examination, for which 329 (96.8%) can be expected to be correct. The FP rate has been reduced to 11 cases (3.2%).

A total of 85 (17.8%) negative results are expected when retesting the 478 cases that are expected to produce positive results at the initial exam. If negative results at the second examination are used to make truthful classifications following an initial positive test result the effect will be a reduction in overall decision accuracy. Only 51 (60.0%) of 85 negative results are expected to be correct, with 34 (40.0%) FN errors. Of the 478 cases that are expected to produce positive results at the initial exam, 53 (11.1%) cases are expected to produce inconclusive results at the second exam. Of the 425 non-inconclusive results



Table 6. Table summary of Bayesian posterior probabilities for subsequent examination results that concur with the initial test result using the posterior results from an initial examination (shown in Table 3) as the prior probability estimate.

	Guilty(deceptive)	Innocent(truthful)	Totals (expected value)	Correct classifications **	Posterior Probabilities
Positive	.812*.849*478=329(TP)	.144*.151*478=11(FP)	340	329	.968
Negative	.083*.849*478=34(FN)	.717*.151*478=51(TN)	85	51	.600
-	-	-	425	380	.894

* Correct classifications are rounded down to the floor integer. Incorrect classifications are rounded up to the ceiling integer.

at the second examination, 380 (89.4%) can be expected to be correct when retesting only those cases that produced positive results at the first exam.

Positive test results that concur with the initial positive test result can be expected to have greater accuracy than positive results from the initial exam. However, negative test results that do not concur with the initial test results can be expected to have lower accuracy. Retesting cases that produce positive results at an initial exam can be expected to reduce FP errors, but can also increase FN classification errors if negative classifications are made when the results of the second exam do not concur with positive results from the initial exam result.

What Happens When We Retest only When Someone Passes the First Test?

Table 7 shows the expected results when retesting only those cases that produce negative (i.e., truthful) test results. A total of the 345 conclusive results are expected at the second examination, for which (84.1%) are expected to be correct. Of the 259 cases for which an initial negative test result is followed by a subsequent negative test result 256 (98.8%) of the results are expected to be correct with 3 (1.2%) FN errors. For the 86 cases that are expected to produce positive test results when retesting case that had previously produced a negative test result 34 (39.5%) can be expected to be correct along with 52 (60.5%) FP errors.

Table 7. Table summary of Bayesian posterior probabilities for subsequent examinations using the posterior results from an initial examination (shown on Table 2) as the prior probability estimate.

	Guilty(deceptive)	Innocent(truthful)	Totals (expected value)	Correct classifications	Posterior Probabilities
Positive	.812*.105*400=34(TP)	.144*.895*400=52(FP)	86	34	.395
Negative	.083*.105*400=3(FN)	.717*.895*400=256(TN)	259	256	.988
-	-	-	345	290	.841



Of the 400 cases that are expected to produce negative results at the initial exam, 55 (13.8%) can be expected to produce inconclusive results at the subsequent test, and 86 (21.5%) positive results are expected. If positive results at the second examination are used to make deceptive classifications following an initial negative test result, only 34 (39.5%) of 86 positive results are expected to be correct, with 52 (60.5%) FP errors. Of the 345 non-inconclusive results at the second examination, 290 (80.1%) can be expected to be correct when retesting only those cases that produced negative results at the first exam.

Negative test results that concur with the initial negative test result can be expected to have greater accuracy than negative results from the initial exam. However, positive test results that do not concur with the initial test results can be expected to have lower accuracy. Retesting cases that produce negative results at an initial exam can be expected to reduce FN errors, but can also increase false positive classification errors if positive classifications are made when the results of the second exam do not concur with negative results from the initial exam result.

Summary

Results described herein are a mathematical description of the Bayesian posterior probabilities that can be expected using test accuracy characteristics reported by APA (2011) as shown in Table 1. Table 2 shows the pattern of results that can be expected when testing under a relatively low prior probability, also referred to as an *informative prior* or *strong prior* because of the strong effect it has when combined with the information extracted from the test data. As shown in Table 3, different prior probabilities will lead to different posterior probabilities. Tests are often conducted under a uniform or *weak prior* probability when there is little information available to describe our knowledge before the test. Test accuracy characteristics under the uniform prior are not different from the available descriptive information that summarizes test sensitivity, specificity, FP and FN errors, and the resulting level precision that can be achieved by the test. A uniform prior is not necessarily uninformative since it is equiva-

lent to interviewing one of two suspects, one of which must be guilty.

What Can Be Learned About Prior and Posterior Probabilities of Polygraph Test Results?

Firstly, observed accuracy of tests conducted when the prior probability is low can be expected to differ from the observed accuracy under the uniform prior probability. It can be expected that the proportion of FP results can exceed the proportion of TP results whenever the expected FP rate exceeds the prior probability or incidence rate. To ensure outcomes for which the TP frequency exceeds the FP frequency it may be useful to select test target issues for which the prior probability or incidence rate exceeds the test FP rate. Alternatively, selecting a decision threshold that constrains the FP error rate to a level less than the prior probability can also ensure a greater proportion of TP than FP results. As a general principle, the FP rate will tend to be high whenever the prior probability is low.

Results shown in Table 3 show that the use of a uniform prior probability of 50% can be expected to yield results that are similar those obtained through frequentist inference. Table 4 shows the improvement in the posterior probability and decision accuracy that can be observed when the results of two examinations concur. Decision accuracy in this example exceeded 97% for concordant results.

Table 5 shows that overall decision accuracy for subsequent polygraphs can be expected to be in the expected range of accuracy for a single examination when classifications are based on all results from the second examination regardless of whether they concur with the results from the first test. This is not surprising, because consideration of all test results from the second examination, regardless of whether they concur with the first examination is, for practical purposes relying singularly only on the results of the second examination. There is virtually no increase in overall test accuracy unless the analysis is limited to subsequent test results that concur with the initial test result, though the proportions of FP and FN errors can be expected to change. This may have important implications for polygraph screening programs in that sub-



stantial increases in test precision may be achieved only when the results concur across successive examinations. Another important practical consideration is that conducting two polygraph examinations on every person can be expected to effectively double the cost and time of testing.

The expected pattern of posterior probabilities when re-testing a case that previously produced an inconclusive test result is also shown in Table 5. These results show that more FP than FN errors can be when re-testing inconclusive cases, though decision accuracy can be expected to be greater for negative results than for positive results.

Table 6 shows the pattern of results that can be expected when retesting only those cases that produce positive results at the initial examination. Accuracy of positive classifications can be expected to increase when these are made for only those cases for which the two tests produce concordant positive results. Retesting positive results can also be expected to reduce FP errors, but only when the results concur for the first and second exam. The FP error rate in this example was 3.2%. Of concern to some will be the fact that FN errors may be increased (to 40.0% in this example) when negative results, following an initial positive test results, were used to classify the test result.

This has important implications for target selection in polygraph screening programs, in that it may be useful to identify and investigate higher base rate behaviors, as the resulting posterior probability estimates can have greater precision. Investigation of higher base-rate behaviors will also allow for intervention and risk management efforts at the level of smaller and more common problems that are more easily resolved. This will be in contrast to attempting to resolve high cost, low incidence problems, for which prior probability estimates can lead to lower precision of the test result. Depending on the relationship risk level and economic cost per incident, and if rare but high risk activity can be characterized as a progression from low risk activities, then

effective risk management of lower risk problems may reduce the incidence rate further for more rare and costly high risk problems that may be more difficult or impossible to fully correct.

The pattern of results that can be expected when retesting only negative test results is shown in Table 7. Accuracy of negative classifications was shown to increase when these are made only for those cases for which the results of the two exams concur. Retesting negative test results can be expected to reduce FN errors. In this example the FN error rate was reduced to 1.2%. However, decision accuracy for positive results, when retesting only those case for which first test result was negative, was only 39.5%.

To summarize:

- When the prior probability is low the false positive FP error rate will be high.
- Overall decision accuracy will be increased only when classification of deception or truth-telling is based on two concordant examination results.
- Retesting deceptive results will reduce FP errors when the results concur, but does not increase overall accuracy when results do not concur.
- Retesting truthful results will reduce FN errors when the results concur, but does not increase overall accuracy when results do not concur.
- Negative results following a positive result on the same examinee will have a higher probability of FN error.
- Positive results following a negative results will have a higher probability of FP error.
- Retesting previously inconclusive



results can provide overall accuracy similar to a single exam, though truthful classifications may be somewhat more accurate than deceptive classifications.

What Are the Limitations of this Analysis?

The most obvious limitation to this analysis is that it is limited to a mathematical example, though mathematical examples can be highly useful and informative because they require us to declare our knowledge and assumptions and then allow us to focus solely on the probability problems themselves. In this analysis, all testing errors are regarded as random variables, meaning that each case is arbitrarily assumed to have the same potential for error as all other cases. If there exists some systematic cause or causes - in terms of examinee psychology, physiology, or sophisticated subterfuge - that contribute to testing error for some cases, then those same causes might be present for those cases at both the first and second polygraph test. In the analytic contexts, systematic variables are often modeled and studied as random variables¹⁷. It is reasonable to assume that both systematic and random errors are included in the available data that anchor our existing knowledge base for polygraph test accuracy. It will be important to continue to study and analyze the effects of repeated testing and to compare this analysis with laboratory and field sampling data.

Another limitation of this analysis is that these examples have addressed repeated testing with a sample of cases. This analysis has not demonstrated the computationally similar practice of computing posterior results for repeated testing of an individual examinee. At the individual case level, a decision to conduct a second or subsequent examination, if influenced by the results of the initial examination, is not independent of the test result. This is similar to confessions that are obtained

as a result of additional questioning that is prompted by a polygraph test result. This limitation could be avoided by re-testing every examinee regardless of the test result.

There may be economic and practical costs to field practice policies that involve conducting two examinations on every person. It may be possible to reduce economic and resource demands by identifying and selecting high-value or high-interest cases for re-examination regardless of the initial test result. Regardless of whether the initial test result is or is not the causal condition for a second polygraph test, a second test on the same examinee will suffer from a second, more obvious, non-independence limitation. Multiple examinations on the same examinee can be expected to have a shared source of response variance: the examinee. If there is some systematic cause, in terms of examinee psychology, physiology, or sophisticated subterfuge, that would cause a polygraph examination to be incorrect then it is possible that the same condition could exist for a subsequent polygraph test.

Finally, this analysis includes no utility function. A utility function could be used to compute a preferred conclusion that maximizes an expected utility objective (e.g., reduced economic costs associated with operational disruption) with respect to the combination of the conditional posterior probability and the pragmatic costs associated with different types of classification error. At present, utility value is managed either heuristically or intuitively or conveniently through field practice policies and program policies that reflect operational objectives, though not through quantification and analysis of the economic or practical value of polygraph test results or costs associated with testing errors. Use of a mathematical or probabilistic utility function could help to reduce the influence of subjective decision making and might also illustrate the potential costs and benefits associated with application of different polygraph policy and practice solutions.

¹⁷ For example: traffic patterns can be modeled and studied as an interaction of random variables including changes in changes vehicle density, velocity and direction, along with other factors, though in reality, the action or behavior or individual cases (vehicles) is not random and is determined by systematic causes related to time, destination and task objectives.



What Are the Limitations Associated with Bayesian Statistics?

One limitation of Bayesian analysis is that it requires the explicit declaration of an input parameter that is often unknown – the prior probability. Sometimes very little is known about the prior. Prior probability estimates can seem to be subjective or arbitrary at times when there is little information available. Although a uniform prior can be used to reduce subjectivity and achieve a more objective outcome, prior probabilities, in field settings, can vary quite widely from 50%. When available information can only bluntly imply a low or high prior probability, then a *Subjective-Bayes* approach might set the prior probability at 90%, 10%, 75% 25%, or any other acceptable and useful prior probability estimate. However, there may be objection to the use of subjective prior information that can have a strong effect on outcomes. Most importantly, different prior probabilities can lead to different posteriori probabilities. *Empirical-Bayes* methods can be used to estimate the prior probability distribution if data are available. It is also possible, when doing Bayesian analysis, to evaluate the observed data against either a central estimate of the prior probability, or against the upper or lower limit of the estimated range of possible prior probabilities. Data can also be evaluated against a range of prior probabilities. The Information Gained Index (Handler, Honts & Nelson, 2013; Honts & Schweinle, 2009; Wells & Olson, 2002) is an example of such an approach, when the precision of a test is evaluated against a range of possible prior probabilities.

A requirement for the specification of an input parameter is not unique to Bayesian analysis. Both Bayesian and frequentist statistics require the explicit declaration of some *a priori* information: either an assumed prior probability or a tolerance for error. Similarly, both Bayesian and frequentist statistics prompt the specification in advance of what are the different possible hypothesis or conclusions that might be supported by the data or evidence from a test or experiment. Both Bayesian and frequentist results can be de-

scribed in terms of probabilities that also can be interpreted categorically according to specified rules. As mentioned previously, Bayesian and frequentist statistical methods can lead to similar results under some circumstances.

Another limitation or criticism of Bayesian analysis is that in assigning a prior probability value it treats hypothesis or conclusions as probability events when, in reality, they are either correct or incorrect. The same criticism can be leveled against the assignment of conditional probability values for deception and truth-telling. Philosophical subtleties notwithstanding, statements are either factually true or factually false. Complexity arises herein because of the desire to make scientific (i.e., testable and falsifiable) statements about some phenomena that is not, of itself, available for direct observation of the facts or for linear measurement. Quantification of amorphous phenomena is an inherently probabilistic endeavor, and all test results are ultimately probability statements.

A Bayesian approach is explicitly probabilistic in that everything measured is a probability, including whether a hypothesis or conclusion is correct. The prior probability assigned to a hypothesis or conclusion merely expresses, in the *Empirical-Bayes* context, our extant knowledge and evidence about the prior probability or, in the context of *Subjective-Bayes*, our degree of subjective confidence that the hypothesis or conclusion is correct.

The goal of Bayesian analysis is to update a prior probability conditional on the observed data. Bayesian probability statements about truth and deception are an attempt to answer the following question: what can be reasonably said, based on replicable analysis of available information, about the level of confidence or degree of belief that can be assigned to our conclusion about deception or truth-telling regarding the investigation target issue? Of course, probability statements of any type may have the potential to be less socially and emotionally impressive than more subjective statements that convey a false sense of certainty. But probability statements are in-



herently more reproducible and realistic than statements based on persona and bravado if they are based on sound testing theory and sound mathematics.

What Potential Hazards Exist with Repeated Polygraph Testing?

Among the most obvious of hazards, when repeating a polygraph test or re-testing an examinee, will be the simplistic error of attempting to assume that the results of one polygraph can be taken to confirm or validate the results of another polygraph. Results of a second polygraph test cannot be taken to confirm a previous polygraph result because two polygraph tests on a single individual, conducted with the same basic methodology and instrumentation, may be expected have the same vulnerabilities to potential error for that individual. Confirmation of a polygraph test result must be independent of the polygraph test result. More broadly, test validity is a broader set of scientific concerns that is not demonstrated at the level of the individual case. [See Nelson (2016) for a brief description of different types of validity and their practical application to the polygraph context.]

Criterion validity or predictive validity, referring to how effectively a test classifies or predicts the phenomena of interest, is a scientific matter involving both research and replication. For practical purposes validity of a test can be demonstrated when it satisfies some basic requirements: a) the testing procedure makes correct use of some theory about how the universe and reality work, b) the testing procedure results in correct classification and prediction at rates greater than can be achieved by chance, and c) probabilistic inferences produce probability coverage estimates that are reasonably close to what is observed in real-world data. Validity of the polygraph test will therefore be a matter of: a) whether the hypothesis is correct or incorrect that recordable differences occur in physiological activity in response to different types of test stimuli, b) whether the testing procedure can actually record and make use of those difference to compute a probability estimate that can be used to make more effective classifications of deception and truth-telling than could be achieved by chance, and c) whether the estimated proportion of correct and incor-

rect deceptive and truthful classification can be shown, over time, to be reasonably close to that observed in real world data.

A similar hazard will be to assume that a confession can *confirm* a polygraph test result. This is incorrect because confirmation of a scientific test result must be based on information that is independent of the test result. Confessions obtained following a polygraph test are, in part, dependent on the polygraph test result if the decision to pursue a confession is triggered by the polygraph test result. Again, confirmation of a polygraph test result must be independent of the polygraph test result. In actuality, it always occurs that some tests results are correct, and some are incorrect. The issue of concern here is the probabilistic margin of error or level of confidence that can be attributed to the test result.

The limitations associated with making attributions about test effectiveness based primarily on actions that are prompted by the test result becomes more obvious when considering what happens when only positive test results are subject to further action. In that case FN errors may be less likely to be confirmed because the examinee may not be subject to further questioning and additional testing. FP errors will be decreased when the two test results concur, with no reduction in FN errors. Instead, if negative results are accepted after the second test there will be an increase in FN errors. The net effect will be a distorted understanding of FP and FN rates. A potential solution towards reducing FN errors might be to subject every examinee to additional questioning after first the polygraph exam regardless of the test result, or to question an individual based on the subjective intuition of an expert examiner, but this would be to reduce the polygraph test to pseudoscientific event – a bogus-pipeline prop (Jones & Sigall, 1971). For practical purposes there is great value in confessions. It is important to remember that the purpose of a confession is not to validate or confirm a test result but to assist in the investigation or risk management efforts of the referring professional or referring agency.



A practical purpose of a scientific test for deception and truth-telling is, in part, to provide information about which individuals are likely to have or not have information to confess. Similarly, results from an initial polygraph inform us of the degree to which the examinee is likely to produce positive or negative results at a second polygraph test. The posterior probability associated with the results of a second polygraph test are conditional on the results of the first polygraph test. It will serve the goals of the polygraph profession, and the goals of the agencies, communities and countries served by the polygraph profession if field examiners become more familiar with and prepared to make use of Bayesian inference.

There may be concern that retesting without additional information will amount to a form of collusion with an examinee's claim that a test result is incorrect. This concern will be perceived as most acute when there is a desire or impulse to naively regard the polygraph test result as infallible. In that case, any acknowledgement of the potential for error might be perceived as a threat to the validity or legitimacy of the polygraph test. These concerns can be reduced substantially by simply remembering that scientific tests are needed and used whenever perfect deterministic solutions (i.e., immune to random chance and immune to human choice or behavior) do not exist, and when direct physical or linear measurement is not possible. Bayesian methods, because they are inherently probabilistic, may help. Concerns about collusion are further reduced or eliminated when decisions to conduct subsequent tests are regulated by field practice policy and not subjective or individualized decision. For practical purposes, to deny or refuse a second examination simply because there is no new information will be to eliminate an opportunity to use Bayes' theorem to optimize operational objectives and reduce FP or FN errors. Caution must be exercised, however, when the results of the two examinations are not concordant.

There is important concern, though little actual evidence, that practice effects might interfere with the effectiveness or precision of

test results when conducting repeated polygraph tests on the same individual. Potential practice effects associated with retesting might include an unknown potential for habituation or sensitization to the test stimuli, with corresponding unknown changes in the potential for FP or FN errors. Ben-Shakhar and Gati (2003) observed within test habituation of electrodermal responses but did not describe its effect on the discrimination of deception and truth-telling. Elaad, & Ben-Shakhar, (1997) had previously reported effective discrimination with up to 12 repetitions of the test stimuli. Shortly thereafter, Ben-Shakhar and Elaad (1999) reported effective discrimination using sequential presentation of 12 different target stimuli. Liebllich, Naftali, Shumueli, and Kugelmass (1974) failed to find any adverse effect of within test habituation, and showed that repeating a question stimulus sequence up to 10 times can reduce classification errors. Similarly, Nakayama and Kizaki (1990) found observable differences in physiological activity in response to test target stimuli using up to nine presentations of a sequence of test stimulus questions, though they also described enhanced differences with the first exposure to the stimuli. Differences in the first presentation of test stimuli were also observed by Grimsley and Yankey (1994).

Most research on habituation has addressed only within-test habituation, though Dollins, Cestaro and Petit (1998) reported that differences in physiological responses to different stimuli persisted when examinees were subject to re-examination after a period of six days. Another limitation to the existing knowledge base on habituation to polygraph test stimuli is that virtually all studies have involved the concealed information test (Lykken, 1959; Krapohl, McCloughan & Senter, 2009) and not the comparison question test. Although the physiological recording sensors and recorded physiological responses are virtually identical for these two different polygraph testing paradigms, the degree to which habituation effects differ for these two testing paradigms remains a matter of conjecture. No publications were found describing sensitiza-



tion effects as a result of repeated testing or repeated presentation of test stimuli.

Repeated testing practices are not uncommon in field practices settings that make use of the comparison question test. Although there is presently no known pattern of predictable problem associated with these practices, continued research is needed in this area. In practice, it remains the responsibility of the examiner at each examination to conduct the polygraph interview, data collection, and test data analysis in an objective manner that will ensure the integrity and effectiveness of the test regardless of the number of previous polygraphs an examinee may have completed. Apart from the effects of psychological or physiological habituation, probability theory informs us that re-testing positive results will reduce FP errors when the test results concur, but will increase FN errors if negative classifications are made following the initial positive test result. Similarly, re-testing negative results will reduce FN errors when the test results concur, but will increase FP errors if positive classifications are made following the initial negative result.

Finally, results from re-examinations under conditions where the examiner is not blind to the initial test result may be at risk for bias or influence. This hazard may be greatest when the second examination is conducted by the same examiner. Risk for retest bias may be exacerbated through the use of polygraph techniques that depend more heavily on the persona and personal involvement of the examiner as these will inherently rely more heavily on subjective judgment at the time of test administration and test data analysis. This potential may be reduced through the use of the more highly standardized directed-lie testing formats, through automation of the test administration, and through automated algorithmic analysis of test data. The superiority of structured algorithmic testing methods over subjective clinical methods has been known since Meehl (1954), and the potential problems with subjective testing and analysis methods have been discussed since Nunnally (1978).

What Does All of this Suggest About Recommended Best Practices for Repeated Polygraph Testing?

Because testing goals and objectives are always formulated in context, and are always influenced by risks, opportunities, resources and practical objectives, there may not be a single recommended best practice for all polygraph programs or all agencies. Agencies that wish to achieve a high level of decision accuracy for both deceptive and truthful outcomes may wish to consider conducting two polygraph examinations on all examinees. Two test results can be combined using Bayesian analysis in which the results of the first exam are used to determine a prior probability when conducting and analyzing the results of the second exam. When the two test results concur, the effect will be a reduction of both FP and FN errors to a low level. Caution must be exercised in addressing results that do not concur, because this will regress the overall error rate to one similar to that for a single examination, with a corresponding imbalance in FP and FN outcomes. If the test results are going to be used as a basis of information for decisions that affect individual rights and liberties then it may be useful to consider retesting as a matter of routine practice.

Agencies that have the primary objective of reducing FP outcomes may want to consider retesting all positive results. When the results of the first and second test concur, the effect will be a reduction of FP errors to a low level. Again, caution must be exercised when addressing results that do not concur as this may increase the rate of FN errors beyond what is desired.

Agencies that are interested in constraining the FN rate to a low level may want to consider retesting negative results before making a negative classification. When the results of the two exams concur the effect will be a reduction of the FN error rate. Once again, caution must be exercised when the two test results do not concur as this may increase the rate of FP errors.



Agencies with a deep applicant pool of highly qualified individuals may find it to be an inefficient use of resources to retest examinees as a matter of routine practice. If the reduction of FP or FN errors is an important objective then re-testing may be helpful, though FN and FP rates might be more efficiently managed through the selection of cutscores that reflect the desired level of significance or tolerance for error. Agencies with limited applicant resources may find it useful to further investigate important details that may aid in risk assessment, risk management and applicant retention.

Polygraphs conducted on high-value or high-interest cases may warrant the need for increased test accuracy, and a multiple testing strategy may be useful. In the event that a program or agency does not routinely retest all examinees this may present a potential problem. If the decision to conduct a second examination is contingent on the results of the first examination then the increase in test accuracy is one-sided (i.e., reduction of either FP or FN results with a corresponding increase in the opposite form of error). If the goal of retesting the high-interest examinee is to achieve high accuracy regardless of the test result, then it will be important that the decision to conduct a second examination is not contingent on the results of the first exam. A simple way to accomplish this is to identify high-value or high-interest cases in advance and make a decision *a priori* to conduct two examinations. In this way, potential increases in test accuracy can be enjoyed regardless of the initial test result. It may also be possible to achieve testing goals with high-interest or high-value cases using a single-stage testing approach while adjusting decision cut-points to correspond with a required tolerance for error.

Until such time as the polygraph test is less subjective and more fully automated in terms of both test administration and test data analysis, there will be the potential that an examiner may permit the results of the first examination to influence the administration or analysis of a subsequent test. For this reason,

program administrators may want to consider various policy alternatives re-testing the same examinee. The most conservative field practice policies would be to require that re-examinations are conducted by a different polygraph examiner.

It will be important for program administrators and policy makers to set clear agency and testing objectives, and to reconcile the benefits and costs associated with various field practice policies. An objective that will be common to most or all polygraph programs will be to make effective use of probabilistic polygraph results. Bayesian analysis offers the advantages of being both easily interpretable and amenable to the combination or aggregation of results from repeated tests. Whereas efforts to teach frequentist inference and caution against misinterpretation have been described as unsuccessful (Sellke, Bayarri & Berger, 2001), probabilistic results from Bayesian analysis, because they are intended to describe directly the effect size associated with the quantity of interest, may be more intuitive for some, and therefore may be less susceptible to misunderstanding and misuse compared to other probability metrics.

Conclusion

Bayes' theorem has become important in virtually every field of science, testing, prediction, and inference including medicine, epidemiology, market research, political polling, forensics, sports, finance, geo-politics and psychometrics. Use of the theorem is computationally simple, requiring only basic algebra. The objective of Bayes' theorem and Bayesian analysis is to update a prior probability to a posterior probability conditional on observed data and evidence that can be represented mathematically and probabilistically. The practical goal of Bayesian analysis is to update and improve the accuracy (i.e., reduce the error or uncertainty) of a probability estimate associated with a conclusion in the form of a prediction or classification.

Bayesian analysis offers the advantage of providing probabilistic results intended to



describe real-world observations. This means that Bayesian probability statements may have more intuitive meaning for some, and are less likely to be misunderstood or misinterpreted by persons unfamiliar with the more complex logic of null-hypothesis significance testing. In other words, results from Bayesian analysis of polygraph test data can be interpreted as a posterior probability of deception or truth-telling.

The posterior probability of deception or truth-telling is the revised probability of a correct classification after taking into consideration the new information from the polygraph test result. Formally, the a posteriori probability is the probability that a case is deceptive or truthful after the test has been conducted and after the test data have been analyzed. Before the test is conducted, and before the test data are analyzed, all available information is summarized in the a priori probability. Bayes' theorem allows us to calculate an updated or improved statistical classifier using both our prior existing knowledge and the known test accuracy characteristics. In the polygraph testing context, a useful advantage of Bayesian updating is that it can be used to combine the results from an earlier polygraph exam with those of a subsequent polygraph exam to achieve a more precise posterior probability estimate about the level of confidence that can be assigned to a deceptive or truthful classification.

Calculation of a statistical classifier can provide us with information that is more likely to be consistent with observable reality in terms of both correct and incorrect test results, and are more likely to be reproducible, compared with merely guessing at the accuracy of the test. Calculation of Bayesian estimates of polygraph test accuracy can help polygraph field professionals, program administrators, legislators and the community to become more thoughtful and realistic about the probabilistic nature of all scientific test results, and to be more accountable for the kinds of decisions that may be informed by the test results.

Reproducible probability estimates have the potential to inoculate polygraph professionals and others against a naïve and frustrating impulse to pretend or expect perfect deterministic accuracy that is somehow unaffected by both random variability and human behavior. Familiarity with probability theory and Bayesian analysis may insulate against subsequent accusations of charlatanism when it is inevitably discovered that testing errors can and do occur, simply because infallible tests of any kind do not exist. Probabilistic test results remind us that scientific tests are not expected to be perfect: they are expected to quantify the margin of uncertainty surrounding a conclusion.

Finally, it is important to remember that the polygraph test is not intended to measure or detect deception or truth-telling *per se*. The polygraph test is merely a tool for the discrimination of deception and truth-telling. Neither the polygraph test, nor any scientific test expected to be infallible. All tests are probabilistic measurements intended to quantify some amorphous phenomena that cannot be subject to perfect deterministic observation or direct physical measurement. All test results are inherently probabilistic and tests are expected only to quantify the margin of uncertainty surrounding a test result or conclusion. Bayesian inference can be useful in the polygraph context because Bayesian posterior probabilities are reproducible evidence-based statements that can be easily interpreted as to the degree of confidence or probability that can be assigned to conclusions that an examinee has been deceptive or truthful.



References

- American Society for Testing and Materials (2005). E2324-04 *Standard Guide for PDD Paired Testing*. ASTM International.
- American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40(4), 196-305. [Electronic version] Retrieved August 20, 2012, from <http://www.polygraph.org/section/research-standards-apa-publications>
- Bayes, T. & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
- Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.
- Ben-Shakhar, G. & Elaad, E. (1999). Effects of questions' repetition on the efficiency of the guilty knowledge test: A reexamination. *Psychophysiology*, 35, S32.
- Ben-Shakhar G. & Gati, I. (2003). The effects of serial positioning and frequency of presentation of common stimulus features on orienting response reinstatement. *Journal of Psychophysiology*, 40, 139-45.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second edition. New York: Springer Verlag.
- Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of P values and evidence. *Journal of the American Statistical Association*, 82, 112-122.
- Bernardo, J. & Smith, A. (1994). *Bayesian Theory*. John Wiley & Sons.
- Box, G. & Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician*, 39 (2), 83–87.
- Casella, G. (1992). Illustrating empirical Bayes methods. *Chemometrics and Intelligent Laboratory Systems*, 16(2),:107-125.
- Casella, G. & Berger, R. L. (1987). Reconciling evidence in the one-sided testing problem (with discussion). *Journal of the American Statistical Association*, 82, 106-111.
- Casella G. & Wells, M. (1993) *Comparing p-values to Neyman-Pearson tests*. Technical Report. Cornell University.
- Carlin, B. P. & Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. Chapman & Hall/CRC.
- Cohen, J. (1994). The earth is round ($p < .05$). *Psychological Bulletin*, 112, 155-159.
- Dollins, A, Cestaro V. & Petit, J. (1998). Efficacy of repeated psychophysiological detection of deception testing. *Journal of Forensic Science*, 43, 1016-1023.
- Elaad, E. & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology*, 34, 587-596.



- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd (Edinburgh).
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Gill, J. (2007). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Second Edition, in the Statistics in the Social and Behavioral Sciences Series., Chapman and Hall/CRC.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Ventari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press.
- Gill, J. (2006). *Essential Mathematics for Political and Social Research*. Cambridge University Press.
- Grimsley D. L. & Yankee, W. J. (1994). The effect of a prior polygraph test on a subsequent polygraph test. *Journal of Psychophysiology*, 18, 13-22.
- Handler, M., Honts, C. & Nelson, R. (2013) Information gain of the Directed Lie Screening Test. *Polygraph* 42(4) 192-202.
- Harris, J., Horner, A. & McQuarrie, D. (2000). *An evaluation of the criteria taught by the department of defense polygraph institute for interpreting polygraph examinations*. Johns Hopkins University, Applied Physics Laboratory. SSD-POR-POR-00-7272.
- Honts, C. & Schweinle, W. (2009). Information gain of psychophysiological detection of deception in forensic and screening settings. *Applied Psychophysiology and Biofeedback*, 34(3), 161-172.
- Honts, C. R. & Peterson, C.F. (1997). *Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer*. Available from the author.
- Hwang, J. T., Casella, G., Robert, C., Wells, M. T., & Farrell, R. H. (1992). Estimation of accuracy in testing. *Annals of Statistics*. 20, 868-881.
- Jackman S. (2000). Estimation and inference via bayesian simulation: An introduction to Markov Chain Monte Carlo. *American Journal of Political Science*, 44(2), 375-404.
- Jackman, S. 2000. Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation. *Political Analysis*, 8(4), 307-332.
- Jones, E., Sigall, H. (1971). The bogus pipeline: a new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76(5), 349-364.
- Iacono, W. G. (2008). Accuracy of polygraph techniques: problems using confession to determine ground truth. *Physiology and Behavior*, 95, 24-26.
- Jeffreys, H. (1939). *Theory of Probability*. 3rd edition, Clarendon Press, Oxford, 1961.
- Jeffreys, H. (1980). Some general points in probability theory. In A. Zellner (Ed.), *Bayesian Analysis in Econometrics and Statistics*. Amsterdam: North-Holland.
- Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, 110, 19313-19317.
- Kass, R. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(431), 1343-1370.
- Kircher, J. C., Kristjansson, S. D., Gardner, M. K. & Webb, A. (2005). *Human and computer decision-making in the psychophysiological detection of deception*. University of Utah.



- Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.
- Krapohl, D. J., McCloughan, J. B., & Senter, S. M. (2009). How to use the Concealed Information Test. *Polygraph*, 38(1), 34-49.
- Krapohl, D. J. & Stern, B. A. (2003). Principles of multiple-issue polygraph screening: A model for applicant, post-conviction offender, and counterintelligence testing. *Polygraph*, 32, 201-210.
- Laplace, S. P. (1812). *Théorie analytique des probabilités*. Paris: Courier.
- Lieblich, I., Naftali, G., Shumueli, J. & Kugelmass, S. (1974). Efficiency of GSR detection of information with repeated presentation of series of stimuli in two motivational states. *Journal of Applied Psychology*, 59, 113-115.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43(6), 385-388.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. University of Minnesota Press.
- Meehl, P. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103-115.
- Meehl, P. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Meehl, P. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical decisions. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What If There Were No Significance Tests?* Mahwah, NJ: Erlbaum.
- Nakayama, M. & Kizaki, H. (1990). Usefulness of the repeated presentation of questions on the psychophysiological detection of deception. *Japanese Journal of Psychology*, 60, 390-393.
- Nelson, R. (2015a). Polygraph testing as a single-subject scientific experiment. *APA Magazine*, 48(5), 101-106.
- Nelson, R. (2015) Scientific basis for polygraph testing. *Polygraph*, 44(1), 28-61.
- Neyman, J. and Pearson, E.S. (1928). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. *Biometrika* 20A, 175-240.
- Neyman, J. & Pearson, E. (1933). On the problem of the most efficient test of statistical hypothesis. *Philosophical Transaction of the Royal Society of London—Series A*, 231, 289-337.
- Nunnally, J. C. (1978). *Psychometric Theory*. Second edition. McGraw-Hill.
- Office of Technology Assessment (1983). *The validity of polygraph testing: A research review and evaluation*. [Re-printed in *Polygraph*, 12, 198-319.].
- Pericchi, L. & Walley, P. (1991). Robust bayesian credible intervals and prior ignorance . *International Statistics Review* 58(1), 1-23.
- Podlesny, J. A. & Truslow, C.M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78, 788-797.



- Poper, Karl. (1959) *The Logic of Scientific Discovery*. First English edition, Hutchinson & Co.
- Raskin, D., Kircher, J. C., Honts, C. R. & Horowitz, S.W. (1988). *A study of the validity of polygraph examinations in criminal investigations*. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc. Boston, MA. URL <http://www.rstudio.com/>.
- Ryan, A. M., Sacco, J. M., McFarland, L. A. & Kriska, S.D. (2000). Applicant self-selection: correlates of withdrawal from a multiple hurdle process. *The Journal of Applied Psychology*, 85, 163-79.
- Sellke, Bayarri, M. J. & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *American Statistician*, 55(1), 62-71.
- Senter, S., Weatherman, D., Krapohl, D., & Horvath, F. (2010). Psychological set or differential salience: a proposal for reconciling theory and terminology in polygraph testing. *Polygraph* 39(2) 109-117.
- Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, 8, 155-167.
- Western, B. & Simon J. (1994). Bayesian inference for comparative research. *American Political Science Review*, 88(2), 412-423.
- Winkler, R. L. (1972). *An Introduction to Bayesian Inference and Decision*. Holt McDougal.



Appendix A.

R Function to Compute a Bayesian Posterior Probability.

```

BayesPost <- function(sens,spec,fp,fn,prior=.5,inv=FALSE,post="Positive") {
  # function to compute the Bayesian posterior probability
  # raymond.nelson@gmail.com (12/19/2015)
  # Bayes' theorem
  #  $P(A|B) = (P(A)*P(B|A)) / P(B)$ 
  # Bayes' theorem written as a hypothesis
  #  $P(H|D) = (P(H)*P(D|H)) / P(D)$ 
  # input sens is the test sensitivity rate
  # input spec is the test specificity rate
  # input fp is the false positive error rate
  # input fn is the false negative error rate
  # input prior is the assumed prior probability
  # inv=FALSE will return the posterior probability of a false positive or false negative
  # inv=TRUE will return the posterior estimate of a true positive or true negative
  # post="Positive" will return the posterior probabilities for positive results
  # post="Negative" will return the posterior probabilities for negative results
  # output is the posterior probability of a true or false positive or negative classification
  ifelse (post=="Negative",
    post <- ifelse( inv==TRUE,
      (prior*spec) / ( (prior*spec) + ((1-prior)*fn) ),
      ( (1-prior)*fn ) / ( ((1-prior)*fn) + (prior*spec) )
    ),
    post <- ifelse( inv==TRUE,
      (prior*sens) / ( (prior*sens) + ((1-prior)*fp) ),
      ( (1-prior)*fp ) / ( ((1-prior)*fp) + (prior*sens) )
    )
  )
  return(post)
}

```

```

BayesPost(sens=.95, spec=.95, fp=.05, fn=.05, prior=.05, inv=TRUE, post="Positive")

```



