

Well done! Or how to Avoid Dangers of Pseudoscience: Common Standard for Research in Behavioural Analysis and Deception Detection in Aviation Security

Jenny K. Krüger¹, María C. Feijoo-Fernández², and Signe M. Ghelfi³

¹German Federal Police Headquarters, Germany; ²Guardia Civil Madrid Airport Unit, Ministry of Interior, Spain; ³Swiss Police Institute, Neuchâtel, Switzerland

ARTICLE INFO

Article history:

Received 4 May 2023

Accepted 22 May 2024

Available online 4 July 2024

Keywords:

Deception detection
Behaviour analysis
Aviation security
Methodology guidelines
Research planning
Experiments in psychology

ABSTRACT

The detection of deception poses one of the main challenges in policing and security environment. It is the inherent goal of security to detect and prevent unlawful events to happen. This is especially true for aviation security as airports continue to constitute attractive targets for terrorist attacks. In consequence, law enforcement agencies are seeking effective and efficient solutions for ensuring high-level security and are often adopting approaches that include behaviour detection. This pressing need for solutions provides ground for pseudoscientific suggestions and methods as those that are cited in an article of the References section. Despite this justified criticism, options to overcome the dangers of pseudoscience are not offered. Therefore, this paper provides a first common standard for conducting research in aviation security for scientists and for practitioners. It highlights several factors that are important to consider before conducting research on behaviour detection. Furthermore, this paper aims to empower experts in the field of aviation security to recognize valid and reliable solutions (e.g., programs, methods, tools) and discusses the relevance as well as the challenges of conducting applied research in the field of aviation security.

¡Bien hecho! O cómo evitar los peligros de la pseudociencia: el criterio común para la investigación del análisis del comportamiento y la detección del engaño en la seguridad aeroportuaria

RESUMEN

Detectar el engaño es uno de los mayores retos los ámbitos policial y de seguridad. El objetivo implícito de la seguridad es detectar actividades ilícitas y evitar que sucedan. Esto es especialmente cierto en seguridad aeroportuaria, ya que los aeropuertos siguen siendo objetivos atractivos para la comisión de ataques terroristas. En consecuencia, los organismos encargados de hacer cumplir la ley buscan soluciones eficaces y eficientes que garanticen un nivel elevado de seguridad y a menudo adoptan enfoques que incorporan la detección del comportamiento. Esta necesidad apremiante de soluciones da pie a propuestas y métodos pseudocientíficos, como los citados en un artículo de la bibliografía de este artículo. A pesar de esta crítica justificada, no se ofrecen opciones para superar los peligros de la pseudociencia. Por lo tanto, este artículo proporciona un primer criterio común para realizar investigaciones en seguridad aeroportuaria dirigido a científicos y profesionales. Se destacan diversos factores importantes a considerar antes de realizar una investigación sobre la detección del comportamiento. Además, el trabajo tiene como objetivo capacitar a los expertos en el campo de la seguridad aeroportuaria en la detección de soluciones válidas y fiables (por ejemplo, programas, métodos, herramientas) y analizar la importancia y el reto que supone realizar investigaciones aplicadas en este campo de seguridad aeroportuaria.

Palabras clave:

Detección de engaño
Análisis de conducta
Seguridad aeroportuaria
Guía metodológica
Planificación de la investigación
Experimentos en psicología

Fortunately, over the last few years, terrorist attacks against airports have been rare (Li et al., 2021). Nevertheless, the threat of terrorism has not disappeared but its form and appearance have changed (Szymankiewicz, 2022). According to the [Swiss Federal Intelligence](#)

[Service \(2023\)](#), the terrorist threat has become more diffuse as individuals act more autonomously and have less and less direct links to al-Qaeda or the “Islamic State”. Furthermore, the jihadist motivated terrorism is not the only form of terrorism, as shown by the tragic

Cite this article as: Krüger, J. K., Feijoo-Fernández, M. C., & Ghelfi, S. M. (2024). Well done! Or how to avoid dangers of pseudoscience: Common standard for research in behavioural analysis and deception detection in aviation security. *Anuario de Psicología Jurídica*. Ahead of print. <https://doi.org/10.5093/apj2024a9>

Correspondence: jenny.dr.krueger@polizei.bund.de (J. K. Krüger).

events of the 2019 mosque shooting in Christchurch, New Zealand (Crothers & O'Brien, 2020). This change in ideological preferences of terrorism is also indicated by the Europol annual Terrorism Situation and Trend Report, that presents developments and key figures of completed, failed, or foiled attacks within the European Union (Europol, 2023). Consequently, strategies against terrorism need to adapt and evolve according to the threat. There is not one-size-fits-all when it comes to counterterrorism. Depending on the context, there are strategies focussing on technical measures (e.g., X-ray screening) or more human based measures such as behaviour detection (BD). However, the evaluation of counterterrorism strategies has revealed that many programs failed their purposes or even increased the likelihood of terrorism occurring (Lum et al., 2006). The evaluation further showed that there is a constant lack of systematic research when it comes to counterterrorism. Regarding BD, there is an ongoing discussion about the validity and reliability behind this method.

In Denault et al. (2020) researchers issue a warning about the dangers of using pseudoscience in security and legal contexts when analysing nonverbal communication pointing to methods like BAI (Behaviour Analysis Interview), programs like SPOT (Screening of Passenger by Observation Techniques), and approaches like Synergology. The authors state that none of them reflect the current state of the science and at the same time they hypothesize possible explanations for why the organizations continue to use these techniques.

As an example of an interrogation tool, BAI is presented by Denault et al. (2020) as an important part of the Reid technique (Inbau et al., 2011). The methodology has its roots in the analysis of nonverbal behaviours to detect what creators believe is a sign of deception or lying, and as a consequence of guilt. Although the creators of this method tried to give support by presenting a scientific study (Horvath et al., 1994), it lacks an accurate methodology, along with the fact that the indicators they tried to demonstrate are not in line with existing scientific research in this field. In spite of this, the Reid technique has a long history and is still one of the most widely taught in a large number of areas (Blair & Kooi, 2004).

Denault et al. (2020) detailed SPOT as an example for a program to identify aviation security threats through the analysis of nonverbal communication. The program has been implemented at various United States airports. The methodology consists of a deployment of Behaviour Detection Officers (BDO), who are responsible for the identification of suspicious behaviours. In order to identify a suspicious behaviour, BDOs are given a list of so-called indicators during their training process (US Government Accountability Office, 2010). Nevertheless, despite the long time that this program has been in place (since 2006), no evidence of its effectiveness has been published or communicated to the public. Throughout this period, the US Government Accountability Office (GAO), along with the Transportation Security Administration (TSA), issued several recommendations to validate the scientific basis of the program (US Government Accountability Office, 2012a, 2012b). In its latest report, GAO holds that the indicators used by the TSA to validate its program are unfounded, so after issuing a number of recommendations, this Office considers that this program should not receive more funds (US Government Accountability Office, 2017). Currently, TSA is in the process of updating this capability by examining methods, protocols, behavioural indicators, and processes based upon recent research in verbal and nonverbal behaviour, the terrorist mindset, as well as experiential information from other BD programs from around the world (J. King Blanchard, personal communication, May 5, 2023). Denault et al. (2020) also present Synergology, that is promoted by its inventors as an approach to read or interpret nonverbal communication. According to the creators of this "discipline" it is based on neuroscience and communication sciences (Synergology, the Official Website, n.d.a.). Again, it is stated that every gesture is

anchored in a mental process, so when teaching to read these signs, students know what the person is thinking or how the person feels at a given moment. As with the SPOT program and the BAI interrogation technique, Synergology has not passed the peer review process that could guarantee reliability and validity.

In trying to explain the use of pseudoscience by some professionals in spite of the fact that none of the above-mentioned programs, methods, and approaches have a solid scientific base, Denault et al. (2020) mention reasons, such as the urgency in solving a problem considered essential (as finding a new airport security measure to detect potential terrorists), a scarce or null knowledge of the scientific methodology and its importance, and the complete underestimation of the real dangers of applying these techniques or even an overestimation of the advantages (misidentifying guilty or innocent individuals by interpreting nonverbal communication).

Denault et al. (2020) focus on pseudoscientific programs, techniques, and approaches and, at the same time, raise what the explanations may be that justify their use. The argument raised by the authors is very important and needs more attention especially in the applied field. However, it would be critical to complete this approach by giving clear guidelines for those who decide to undertake scientific research in the field of behavioural analysis and deception detection as well as for organizations deciding to implement such programs and training.

Goal of this Paper

This paper provides a guideline for conducting scientific research in the field of behavioural analysis and deception detection in order to increase the ecological validity of the conducted research and to foster scientific studies in this field. Although academic diversity in research and various forms of methods and approaches are important, the implementation of general common standards can enhance the quality of research and by this making general (i.e., valid) answers and findings possible. The aim of this paper is twofold, first to complement Denault et al. (2020) article by providing basic standards, and second, to be used as a guide to evaluate the right study design a priori. It is addressed to people linked to the field of civil aviation and policing in some way, such as managers, regulators, legislators, law enforcement officers, practitioners etc., as well as academics.

Throughout this paper, we briefly detail the scarce scientific research conducted in the civil aviation context followed by a list of factors that could explain why pseudoscience is adopted in some cases. We also briefly address the theoretical background in the field of deception detection followed by the description of methodological issues that we must tackle when conducting scientific research. Next, we focus on the research design addressing some key aspects on how to conduct studies in this field, ending with some general conclusions.

Scientific Research Meeting Common Standards within Civil Aviation Context

The 2001 terrorist attacks perpetrated against various infrastructures in the United States are still on everyone's retina. For the first time, a commercial airplane became itself a weapon (Jenkins, 2021). This event generated an urgent need for changes and improvements in security measures as mentioned in Denault et al. (2020). This urgency to find a tool capable of detecting potential threats against civil aviation opened new horizons such as behaviour detection and deception detection in airports. While BD programs began to be implemented, criticism concurrently appeared (academics, governments, and general public) (Blandón-Gitlin et al., 2014; US Government Accountability Office, 2017). Sometimes

entities involved in decision-making tend to act under the rule of thumb of “something is better than nothing” and do not always understand that scientific research takes time. However, we can find some exceptions, like the ones we briefly present here, showing that collaborative research is always possible.

Within civil aviation, BD procedure follows three main steps (UK National Protective Security Authority, 2023): baselining (continuous environmental assessment of the observable behaviour in a given area or context), behaviour observation (to detect behaviours that differ from the baseline), and resolution process (to establish credibility/deception, e.g., by talking to the person, conducting interviews, consequently including a decision whether the person is telling the truth or not). Regarding the latter, decades of research has been conducted. As it is widely addressed in different publications (Vrij et al., 2023, 2022), we will focus this section in the scarce scientific research conducted in the field related to the detection of anomalous behaviour.

A number of countries implement BD programs in airports, but a few carry out scientific research in this context (Denault et al., 2020). Switzerland meets these criteria presenting a program built on scientific results published under peer-reviewed conditions. This research was conducted by a multidisciplinary team integrated by academics, airport police, and investigative police. Analysing Suspicious Persons and Cognitive Training (ASPECT) was initially supported by three empirical studies. First, Koller et al. (2015b) studied how good five different groups were at detecting a thief's intentions (students, police recruits, inexperienced police officers, experienced police officers, and criminal investigators). They find all groups detected thieves before the commission of the theft. Criminal investigators show the best performance followed by experienced police officers, inexperienced police officers, recruits, and students. The main limitation refers to the use of just one type of crime (CCTV footage of thefts). Following up these results, Koller et al. (2015a) focused on specific nonverbal behaviours that can predict a criminal act (moving patterns, communication behaviours, self-adaptors, and object-adaptors). The results showed that offenders display different moving patterns than non-offenders, offenders' communication behaviours differ from the rest of airport users, and offenders use more self-adaptors and less object adaptors. Among the limitations are that the use of self-adaptors may be due to the increase of arousal in stress situations and the number of recordings used in the study.

Spanish Guardia Civil has also conducted scientific research in the field to give support to BD programmes within civil aviation environment. Again, a multidisciplinary team designed and conducted research in the field (university academics, airport police, and police from the criminal behaviour analyst's branch). Feijoo-Fernández et al. (2023) designed and conducted research in a major international airport in Spain. The authors propose a theoretical framework in trying to explain the anomalous behaviours displayed by some airport users and the first definition for anomalous behaviour is included. In a first phase, police officers with experience in the airport environment collected behaviours that preceded any kind of crime. The final list of thirteen behaviours was grouped in patterns of movement, patterns of communication, indicators coming from autonomous nervous system, and object-adaptors. In a second phase, this list of anomalous behaviours was tested in the same context. The results showed significant differences in movement and communication patterns in users who commit illegal activities and users who do not. People who display these anomalous behaviours were more likely to be linked with illegal activities. No significant differences were found for indicators related to physiological changes nor for object adaptors. The most important limitation here is that in a real scenario with real passengers unknown variables could have influenced the sample and although the whole sample was checked, authors could not rule out the commission of illegal activities among those classified as negatives (no crime/offense).

Once detailed the scarce studies conducted in the field to address behavioural analysis to detect anomalies, it seems clear that more research is needed. To find explanations for this scarcity and the use of other methods, in the next section we will detail some of the factors that make pseudoscience possible.

Factors that Make Pseudoscience in Security Context Possible

Some of the factors that can explain the use of pseudoscientific techniques in this area have already been pointed out in Denault et al. (2020). However, the urgency to solve a problem, little knowledge of the scientific method, and the consequences of the use of pseudoscience are some of the possible explanations. For this reason, we expand what we consider to be part of this widespread problem.

Airport security is one of the most legislated fields within civil aviation (Yadav & Nikraz, 2014). In the last two decades, the regulations in security have become extremely complicated in this search for a solution against terrorist threats. However, it is usually the regulators who decide to implement new measures and, as a general rule, they are not specifically trained in academic research. It is very common that these practitioners in charge lack the appropriate training to judge, check academic outputs, and answer the question of whether the offered results are valid and beneficial. In some cases, the results presented to validate the use of a new tool in the field have many limitations among other factors, because only few researchers focus on security and BD research. Needless to say, conducting applied research is very challenging and not as easy to conduct as accurate laboratory studies but field approaches are often needed. Due to the variables that can interfere in this type of research (extraneous or unknown) (Maner, 2016), the results of applied science are even harder to publish than typical academic results. Furthermore, it is very common that field research is classified so the exchange of confidential information on a bilateral basis is complicated and sharing information with external researchers is not possible.

Behaviour detection as a topic is kind of “sexy”. Based on personal experience by the authors as well as on the vast amount of available so-called BD trainings almost everybody means to understand its goal, procedure, and background. As a counterterrorism technique it is considered acceptable within the civil aviation field as it is implemented in the guidelines by the International Civil Aviation Organization (ICAO, 2017) since 2017. In contrast to racial profiling, BD only focuses on behaviour and on deviations in behaviour, consequently. Further, as stated by Mineta Transportation Institute, even a straightforward campaign as “If you see something, say something” has significant impact on the prevention of terror attacks (Jenkins & Butterworth, 2018). But the general need for easy, applicable, available solutions in the fight against terrorism (in particular) and criminality (in general) makes people turn to pseudoscientific techniques that offer easy and understandable results often available through internet, books, etc., ignoring that scientific research is not black and white. After reviewing the scarce scientific research done in the civil aviation environment and analysing some of the factors that contribute to the adoption of pseudoscience, we open the next section of this paper by reviewing the theoretical background regarding BD and covering essential methodological issues like research questions and hypothesis generation, sample sizes, operationalization of variables, and cover stories. We will conclude with a section dedicated to some aspects of research design and how to properly conduct research studies.

Theoretical Background

Within civil aviation context, BD can be divided into three main tasks: baselining, observation of indicators, and resolution conversations. Taking these three basic pillars into account, (applied

and laboratory) research insights can provide a solid base to build programs within this environment.

Existing literature offer deep insight and much information regarding the investigation of lies and deception (e.g., [Docan-Morgan, 2019](#); [Harrigan et al., 2008](#); [Vrij, 2008](#)). The following concepts are often covered:

- Baselineing ([Bogaard, Meijer, et al., 2022](#); [Bogaard, Nußbaum, et al., 2022](#); [Ewens et al., 2014](#); [Vrij, 2016](#))
- Cognitive load theory ([Muñoz García et al., 2023](#))
- Criminological theories ([Piquero, 2016](#))
- Heuristics, Bias and Decision-making ([Meterko & Cooper, 2022](#); [Neal et al., 2022](#))
- Interview techniques ([Harvey et al., 2018](#); [Nahari & Nisin, 2019](#); [Walczyk et al., 2016](#))
- Misconceptions on verbal and non-verbal signals ([Denault, 2020](#); [Denault & Zloteanu, 2022](#))

One of the most supported premises in this context consists of cognitive based theories stating that lying is cognitively more demanding than telling the truth ([Muñoz García et al., 2023](#); [Vrij et al., 2022](#)). This assumption is based on the fact that lying is more demanding in terms of executive functions, for instance when suppressing the truth, retrieving important information, and building a lie. Different interview techniques rely on cognitive load theories: imposing cognitive load to impair liars' cognitive resources (e.g., reverse order, keep eye contact, turn taking) ([Mann et al., 2012](#); [Vernham et al., 2014](#)) presenting a model statement or asking unexpected questions ([Porter et al., 2021](#); [Shaw et al., 2013](#)). Nevertheless, some precautions must be taken into account because the use of such techniques with some people can result in an overloaded cognitive capacity and as a consequence it can lead to a misunderstanding of the indicators and false positives.

Human behaviour, including the judgement of information and the subsequent decision-making process, is often not completely rational and tends to be biased under certain conditions. This is particularly true for situations where people's resources are low—for instance, due to attentional distractions, time pressure, or knowledge gaps and misconceptions. The reason for this behaviour is that humans are susceptible to decision heuristics ([Evans, 2006](#); [Gigerenzer & Gaissmaier, 2011](#); [Kahneman & Tversky, 1979](#); [Tversky & Kahneman, 1974](#)). A heuristic is a mental shortcut that is applied to reach an efficient decision. Heuristics constitute part of the daily life and can be applied consciously or unconsciously. Whether applied on purpose or not, all heuristics neglect part of the information given and violate to some extent the assumptions of subjective expected utility.

Having said that, it is clear that one of the most important parts for building a good BD program is the BDOs training process.

It has been shown that individuals tend to assume facing truthful situations and therefore judge messages as truthful ([Buller & Burgoon, 1996](#); [Vrij, 2008](#)). This truth-bias does also make sense in terms of the overall social rules. Most of the individuals believe that lying is not the norm and that a liar is therefore breaking a rule. Even the liar believes that everyone else, except himself, is more or less telling the truth (compare [König, 2020](#), regarding Kant's Moral Philosophy). These assumptions are important to understand lying and deception. Above this, it has been stated that individuals lie frequently in everyday social interactions (e.g., [Buller & Burgoon, 1996](#)). However, recent research on large-scale groups of participants mainly via self-reports indicate that this common understanding does not entirely reflect the reality. Lying and deceiving is not normally distributed, but rather positively skewed ([Serota & Levine, 2014](#); [Serota et al., 2010](#)). This means that the indicated average number of lies on a daily basis is conducted by a minor group of so-called prolific liars ([Serota et al., 2022](#); [Verigin et al., 2019](#)). In accordance with former assumptions, it is well documented that most of everyday lies belong to the category of white or also called pro-social lies. As the name indicates, the content of these lies is not harmful to others and can be stated as

often within cultural understanding accepted ([Bryant, 2008](#)) and expected. As [Saxe \(1991\)](#) stated "Psychologists, as well as others in society, often use deceptive techniques for the 'social good', and there are a number of conditions under which lying is seen as acceptable..." (p. 409). Based on the given focus of this paper on aviation security research seems the further differentiation of white lies ([Erat & Gneezy, 2012](#)) of less importance.

Individual differences in the tendency to lie due, for example, to personality traits have been discussed ([Hart et al., 2020](#); [Jonason et al., 2014](#); [Roeser et al., 2016](#)). The emotional impact of lying has been in focus of scientific interest and discussions for some time past (e.g., [Buller & Burgoon, 1996](#); [Caldwell-Harris & Ayçiçeği-Dinn, 2009](#); [Ekman, 2009](#)) whereas the higher cognitive load of performing a lie (effectively) has been shown (e.g., [Vrij et al., 2008](#)). The affective negative impact of deceptive behaviour and lying has been shown on a neural basis. [Baumgartner et al. \(2013\)](#) investigated individual differences in regard to anterior insula activation pattern and tendencies to deceive. This brain region is known to be associated in mapping internal bodily states and representing emotional arousal and conscious feelings. A high level of baseline activation of the insula region is related to negative affect and therefore the tendency to avoid aversive emotional situations. The results of [Baumgartner et al.](#) did show that individuals of high-level anterior insula activation tend to lie and deceive less, which is potentially caused by the attempt to avoid the associated negative emotions due to lying. It is important to take these insights and general assumptions about the nature of lying and deception into account in order to investigate deceptive behaviour properly, especially in terms of antisocial lies, which might be related to criminal or terror activities. This type encompasses all kind of deception and lying when harmful (ecological, physical, emotional, etc.) or even lethal consequences are accepted. Moreover, lethal outcomes could even be one of the objectives and not only acceptable consequences. Contrary to prosocial lies, detecting antisocial lies is opposite to this, because these kind of lies are very rare ([Serota & Levine, 2014](#)), therefore unexpected and less or not at all trained. However, it can be assumed that antisocial lies are mostly high-stake lies which involves serious consequences for the deceiver and are therefore easier to identify. Further, BD related to antisocial lies is of major importance for aviation security and thus of major interest in the field of research. Given the fact that not all passengers can be extensively questioned regarding their underlying intentions, nonverbal deception detection offers additional insights.

Without taking into consideration some of the key points in understanding human behaviour, we can only end up with non-effective BD programs. In order to improve the scientific evidence for such BD programs, the key aspects of scientific research should be followed—also in the context of applied research in aviation security.

Once the main aspects of the theoretical background on BD have been seen, in the next section we will take a closer look at some methodological questions that should be addressed to conduct research in this field.

Methodological Issues

Research Questions

The first step of a research process is the definition of research questions ([Lipowski, 2008](#)). They must be well justified and precisely stated to obtain relevant and credible results. Research questions are especially important when there is only limited research available to rely on or when research results are contradictory or ambiguous. For example, there are studies suggesting that the use of object adopters is an indicator for deception (for a discussion see [Koller et al., 2015a](#), whereas other studies could not confirm that; [Feijoo-Fernández et al., 2023](#)). By stating research questions, these findings can be taken

into account. There are three main categories of research questions: “descriptive” questions, to ask whether a phenomenon exists and if it exists to ask about descriptions/classifications or composition; “relational” questions, to ask about the relationship among different phenomena; and “causal” questions, to find out the origin of that relationship (Huber, 2014). A research question is considered as good when it is feasible, interesting, novel, ethical, and relevant.

In the context of civil aviation (detection of anomalous behaviour and lie/truth detection), it is considered essential to define whether the study should focus on:

- General understanding of human deceptive behaviour,
- General understanding of deception detection,
- Verbal or non-verbal behaviour,
- General behavioural patterns or accurate indicators of deception,
- Deceptive behaviour during interview situations or without any interaction (just observation), and/or
- Aspects of implementation and practicability.

Further, it should be evaluated whether a qualitative or quantitative design is more appropriate to answer research questions. For this evaluation, additional factors can be crucial, for example feasibility of conducting an experimental design in the field, scientific expertise, budget and time restrictions, ethical aspects, and even specific policy regulations at the airport (Lowhorn, 2007). Nevertheless, in the case of qualitative research questions, it is advisable to evaluate if certain aspects could be investigated quantitatively (Khaldi, 2017). For one aspect, qualitative research is generally more challenging to publish (Petticrew et al., 2008). Furthermore, from a methodological point of view, due to the fact that research results on deceptive behaviour often do offer a wide range of interpretations, it seems important to assess objective data based on quantitative or qualitative research attempts (e.g., Hauch et al., 2017; Koller et al., 2015b). Hence, before starting the research, it needs to be clarified whether the research question refers only to qualitative aspects (e.g., analysis of police officers' experiences collected through interviews), whether some aspects of the research question can also be measured via quantitative methods (e.g., number of observable indicators, like gestures, instead of interpretations, like nervousness), or whether all interesting facts can be operationalized and measured via quantitative methods (e.g., movement patterns).

To summarize, the goal of research questions is fourfold: Firstly, to boost scientific work, secondly, aim to systematize knowledge, thirdly, to explain the phenomenon of the study, and fourthly, to serve as a link between the knowledge found in the past and what we are seeking to learn in the present. To increase the accuracy of the planned study, hypotheses need to be stated in a next step.

Hypotheses Generation

In general, a hypothesis provides a logical and feasible answer to a problem without knowing whether this is actually true or not. Hypotheses are often generated based on a theory or a literature review. Establishing hypotheses is important for addressing causal research questions. Hypotheses should be clearly written, so they are understood by the field experts and scientists, they should not be contradictive, and every hypothesis should be addressed with a subsequent analysis. Among other characteristics, hypotheses should not contain ambiguous words and propose a relationship between two variables (independent and dependent) (e.g., Howitt & Cramer, 2011; Huber, 2014).

More specifically, a hypothesis should be considered as a tentative explanation for a specific phenomenon, therefore subject to empirical validation. A requirement for a hypothesis is that it can be falsified. Based on the analysis, the hypothesis is either rejected or considered confirmed with high probability. In other words, a hypothesis can never be truly proven right, only falsified. On the other hand,

hypotheses that are rejected contribute to science by enhancing what we know and do not know about a specific phenomenon. For instance, it has been shown that nervousness is usually very high in research in aviation security and, at the same time, not significant because it occurs due to the context, personality, personal situation etc. (e.g., Feijoo-Fernández et al., 2023).

For generating adequate hypotheses it is essential to have very good knowledge of the field, the topic, and the specific circumstances in the area that should be investigated. Therefore, we would like to stress the importance for a close collaboration between scientists, practitioners, and experts in civil aviation in order to develop critical hypothesis and enhancing the scientific knowledge in the field.

Sample

One crucial part in the research planning refers to the sample, its size and composition in relation to the question that should be answered by the study. Therefore, the number of groups and composition of each group should be appropriate in order to answer the given research questions. For example, to investigate how terrorists might behave prior to the attack and how to recognize the attacker is not recommended to instruct students to imagine being a terrorist and try to kill as many innocent bystanders as possible. Although this design offers interesting insights, it is not suitable to investigate deceptive behaviour of terrorists in particular besides ethical concerns that would come along with such an instruction. Undergraduate students of psychology which constitute the sample of many studies, might not be the most adequate group of participants for such a research question. Especially in behavioural research, as in detection deception, the potential impact of specific group characteristics should be considered and the universality of the yielded phenomena questioned (Henrich et al., 2010). Next to the point that students do reflect a certain population who is able to choose an academic training, research insights did pose the question whether differences between students and other adults (Serota et al., 2022) or even in regard to specific areas of academic or professional field do exist (Gerlach & Hertwig, 2019; Verigin et al., 2019). Further to personal and/or external factors, age is an influencing factor which has to be taken into account for research on deceptive behaviour. Given the general neuro-psychological changes as the maturation of the prefrontal cortex, essential executive functions, e.g., critical thinking, decision making or impulse control, need to develop and evolve with age (Diamond, 2018; Stuss, 1992). Juvenile criminal law reflects the fact that juveniles cannot completely accomplish the above-mentioned aspects (e.g., impulse control). Hence, certain factors such as age, gender, culture, just to name a few, have to be considered as critical for the selection of representative participants. In this context, critical impacting factors need to be indicated and evaluated regarding their importance for the future investigation. For example, native language might be of importance for a study including interview situations, whereas it might be less essential for exploring nonverbal behaviour patterns, except for ensuring understanding of given study instructions.

Referring to the very first step, the critical questions that have to be answered are:

- Which groups can be a representative sample for the research question that has to be investigated? and
- What are impacting factors or conflicting aspects in order to compose a reliable sample?

Controlling impacting factors within the psychological research by randomized repeated measurements over the exact same sample is a well-established method (Davis, 2002; Keselman et al., 2001). Although, this procedure could eliminate crucial, unintended influences, for instance differences between individuals regarding

personality traits, in a study on detection of deception it might not be feasible due to methodical contradictions (compare the sections “cover story” and “laboratory or field studies”). But an a priori precise definition of the sample composition combined with a sufficiently large sample size can obviate the influence of unintended factors (Wicherts et al., 2016).

Further, given a precise research question and the statistical procedures that are going to be performed should be approximately defined. On this basis, the researcher is able to calculate the required sample size (Erdfelder et al., 1996). Computer programs, such as *g*-power, can define the minimum sample size in total (and per group) a priori (Faul et al., 2007). Although this step is of major importance, to conduct research that might yield valid, reliable results, it is often missed to conduct or to report (Kyonka, 2019). This simple tool offers the calculation of the precise sample size in consideration with effect size and power, two factors which are crucial for high-level publications and/or application of revealed results (Button et al., 2013). By leaving this step out, researchers miss the opportunity to optimize and adapt their studies in advance. Performing a power analysis a priori demands the researcher to clarify whether for instance a group comparison with one or more factors (e.g., MANOVA) would be the correct method. Since every researcher has to figure out which statistical procedure is correct for the given study at some point, no definition in advance is only a postponement. Although conducting a study offers insight and deeper knowledge about the nature of its data and therefore knowledge about the correct statistical analysis increase. Nonetheless, exact planning prevents from collecting data of, for example, too many participants or of less (important) data (Cohen, 1990; Wicherts et al., 2016).

Operationalization of Variables

Based on the formulated hypotheses, firstly, the dependent and independent variables need to be defined. A variable that is assumed to be a cause is called independent variable. A variable that is assumed to be an effect is called dependent variable. Secondly, the variables have to be operationalized in order to measure them (compare Price et al., 2015). This process refers to the connection of theory to empiricism and can be challenging. Some variables can be assessed directly, for example, detection performance of BDOs. However, other variables such as deceptive behaviour are more complex or abstract and can be assessed in different ways. The operationalization must be valid and as close to the “real” variable as possible. It is recommended to conduct literature research and understand how the variables have been operationalized in published studies (e.g., Price et al., 2015). For some variables there are already validated and reliable operationalization and assessment methods available (e.g., personality traits). Furthermore, it is important to think about the possible interpretation based on the assessed variables. This needs to be taken into account for the answering format. Depending on the measurement level, different types of statistical analysis can be undertaken. For example, if detection performance is only assessed with yes (detected) and no (not detected), it is called a binary variable and only limited analysis is possible. Often, there is merit in measuring more than one dependent variable and to apply different operationalizations, respectively (Wicherts et al., 2016).

Cover Story

Conducting research on BD implies in most cases the use of a cover story. There are multiple reasons for this approach. To demonstrate the importance of an appropriate cover story, the complex nature of lying and impacting factors needs to be understood.

Past research has shown that targeting deceptive behaviour is possible when the stakes are high (e.g., DePaulo et al., 2003; Frank & Ekman, 1997; Frank & Feeley, 2003). Participants can display natural signs of deception if they are motivated to lie.

Hence the cover story, in which the participants have to perform and refer to, needs to be very well created. All cover stories for the different groups (e.g., experimental and active control group) should focus precisely on the targeted research question, the applied methods, as well the specific sample composition. Given that deceptive behaviour (verbal or nonverbal) is the focus of the investigation, the participants receive a task that demands them to lie and to deceive. Via this story participants are engaged in the task they are willing to accomplish even if they have to deceive others.

Therefore, the specific cover story for the investigation needs to ensure:

- High intrinsic motivation of all participants to perform,
- High-stake lies, including reward and punishment within ethical limits,
- Reference to the real world,
- Excluding moral, ethical conflicts, and
- Being indistinguishable as cover story.

Cover stories of high quality encourage the participants to lie without explicit request to do so and without participants' awareness of being in the genuine focus of research interest. Subtle cover stories are explicitly designed for the specific composition, needs, and background of the participants.

In order to investigate deceptive behaviour, it is of major importance that the participants behave in a natural way while lying. It can be stated that deceptive behaviours are based on a subjective intention to deceive rather than an objective state (e.g., Fernández & Halty, 2018; Sip et al., 2008). Even though real-world scenarios cannot be completely imitated in a research design (laboratory or field approach), due to different restrictions (e.g., ethical guideline), the individual importance of the objective to archive needs to be taken into account. Therefore, we believe that only high stake lies, including the announcement of “real”, personally important objectives (reward) and real consequences when failing (punishment), lead to representative behaviour of deception. Due to this, once again, research studies that request participants to imagine being a terrorist do not fulfil these requirements. One can assume that such a study design would investigate the cognitive ability for imagination, behaviour during moral, ethical conflicts, as well as dutifulness and compliant behaviour etc., just to name a few. Without a doubt, these are very interesting aspects for psychological investigations, but not appropriate to understand deceptive behaviour.

Consequently, participants' naivety regarding the study objectives, as well as the main overall research question, is crucial to use a cover story correctly. Only by ensuring that the participants have no clue about the real research question, natural behaviour while relying on high-stake lies can be observed and analysed. Accordingly, it may be impossible to apply repeated measurements in most of the investigations. Although it first seems appropriate to control for impacting factors on deception, for example personality traits such as neuroticism (Hart et al., 2020) or narcissism (Jonason et al., 2014), repeated participation might be impossible due to the necessary naivety. Here, we see how important it is to precisely plan all aspects of the investigation in advance since the main aspects mutually influence one another.

Given this, a pilot study is of importance to ensure that the applied story cannot be identified as a cover story and that the research goals can be achieved by applying that particular legend. Double-checking is crucial, whether the design focuses genuinely on deception, other explanatory alternatives, for example moral conflict can be excluded and relevant possible influencing factors, such as personality traits or surroundings, are somehow under control.

Finally, and of high importance, is to ensure participants' approval. Since investigations with a cover story do not allow for giving all relevant information in advance to the participant in full, precise debriefing including the possibilities to answer all kinds of questions is necessary. Here, planning a cover story has to include the consultation of the legitimate ethical standards and/or requirements of works and staff council. In order to perform research on deceptive behaviour including a cover story ethical guidelines need to be taken into account (e.g., [Howitt & Cramer, 2011](#)). As the World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects states, every participant must be adequately informed in regard to objectives, methods, and potential conflicts of interest amongst others. This includes that, due to the exact research question to be investigated, participants might learn about the genuine methodology and research questions after finalizing the research project (Declaration of Helsinki; [World Medical Association, 2013](#)). Given this, the original agreement for participation can be revoked.

Experimental and Active Control Groups

In terms of understanding causal dependencies, applying a scientific experimental design is the most common and valuable way for research (e.g., [Price et al., 2015](#)). An experiment is defined as manipulating one aspect and investigating the effect on a defined variable (e.g., [Huber, 2014](#); [Hussy et al., 2013](#)). Independent of the exact study design, an experiment at least consists of an experimental group and a control group or two points of measurement (e.g., repeated measurement). In the field of deception detection research, it is especially important that the control group is comparable to the experimental group. This means that the control group should perform an active control task that is comparable to the task of the experimental group (e.g., [Koller et al., 2020](#)). Only by conducting a statistical comparison between the two groups, valid and reliable results can be achieved. To accomplish a comparable control group, the main psychological tasks and processes that the experimental group is performing have to be exactly defined. For instance, if the study design refers to security measures within an airport, both groups need a clear purpose of what they have to do at the airport. The goal of such a study is to detect and distinguish passengers with malintent (experimental group) from normal truth-telling passengers (control group). While the experimental group receives a task that includes the need for deception and/or the task that motivates participants to deceive, the active control group should also receive a specific task in order to imitate the normal passenger behaviour. From a psycho-cognitive perspective, the behaviour of passengers at an airport includes some kind of visual search (e.g., search for gate, security check, or coffee shop etc.; e.g., [Feijoo-Fernández et al., 2023](#); [Weinberger, 2010](#)). Therefore, the task should stimulate the control group to perform this kind of behaviour without requesting them to mimic directly.

Including an active control group makes it possible to:

- Assess the typical behaviour of the experimental group,
- Do so without changing environmental influences,
- Control or exclude for external, e.g., cultural, influences, and
- Ensure that observable, statistically detectable differences are only due to the applied scenario.

The review of literature shows that the inclusion of a control group in the study design is not the common standard and this constitutes one of the major shortcomings of conducted research in deception detection contributing to misbeliefs as well as pseudoscience in this field (e.g., [Denault et al., 2020](#)). It is therefore important to stress once again the relevance of systematically including an active control group in the study design.

Laboratory or Field Studies/within and between Groups Design

One of the next steps is to decide on a laboratory or a field approach. Both approaches have their advantages and disadvantages as well as restrictions. Investigations in laboratory conditions have the advantage that the whole set-up can be planned with precision and can be controlled. Therefore, each and every participant can perform the same task under the exact same conditions, allowing valid and objective data collection. Further, methods can be used that are less applicable in the field, for example physiological measurement of skin conductance. Simulations such as the mock crime scenario can be easily realized in combination with standardized psychological tests, for example, the Concealed Information Test (CIT) (e.g., [Ben-Shakhar, 2012](#); [Koller et al., 2020](#)) and/or assessment of individual differences, e.g., NEO-FFI ([Costa & McCrae, 2008](#)), dark triad ([Jonason & Webster, 2010](#)), or further covert data collections, for instance video recordings ([Koller et al., 2015a](#)). Hence, laboratory investigation can lead to high data quality and therefore to meaningful interpretations of the received results. Unfortunately, the external validity of experiments conducted in laboratory settings is reduced (e.g., [Berkowitz & Donnerstein, 1982](#); [Huber, 2014](#); [Levine, 2017](#)). This means it remains unclear to what extent the results hold true in real life (e.g., [Verschuere & Meijer, 2014](#)).

Consequently, some research questions are better suited for a field experiment. For instance, if the study is investigating aspects of practical implementations (e.g., effectiveness of new procedures), it is often more appropriate to conduct a field study in order to achieve a high external validity (e.g., [Galizzi & Navarro-Martinez, 2019](#); [McDermott, 2011](#); [Taylor & Asmundson, 2008](#)). Although the data collection for a field experiment is highly demanding, it is worthwhile to examine the suitability for the research questions.

A second aspect of the experimental design refers to the comparison of the groups. This can either be between groups or within groups. As mentioned before, an experiment consists of at least two groups that are compared – that is, the experimental group and the control group that are compared (i.e., between groups design) or it consists of two (or more) points of measurement ([Field, 2009](#)). This means that the same group is analysed in different stages (i.e., within-group design or repeated measures). In this case, a field approach might be less feasible, because influencing factors need to be controlled. For instance, the advantage of a within design, such as repeated measurement, is that factors like personality traits, anxiety or fear, that might substantially impact deceptive behaviour or the recognizable signs of a deceptive person, can be controlled within that design. Conducting a within group comparison decreases the total number of required participants and increases the number of controlled influences as well as the validity and reliability of the results. Systematic counterbalanced randomization of the participants to all terms and conditions of manipulations must be guaranteed. However, only a few field approaches fulfil the requirements and necessary circumstances to properly realize a within group design. The randomization of the participants is often not possible in a field setting due to practicality reasons ([Pierce & Balasubramanian, 2015](#)). In a case where the study focuses on deceptive behaviour in comparison to honest individuals under the identical conditions, point in time, surroundings, interaction partners etc. it might not be possible or of any benefit to conduct a within subject design. Knowing the study procedure and sequences of incidents would be an exclusion criterion to participate again in the group for comparison.

The planning aspects to decide for a field or a laboratory design as well as for a between group or a within group design are intertwining. This means, that the process of decision is stepwise and all aspects need to be judged in relation to one another.

Further, it can be possible to implement a mixed design in a large investigation. In this case, the advantages of the different

experimental designs can be combined. Although, the development, preparation, data collection, and analysis for such an investigation are complex and demanding, the advantages outweigh the costs. For example, interesting and novel, valid, and reliable results could be provided by conducting multimodal approaches, such as psychological testing, behavioural observations, and video recordings in the field.

Statistical Analysis

Based on the study design, sample size and statistical procedures have to be planned and calculated before starting data collection. Conducting power analysis a priori reveals the minimum, total sample size related to the statistical analysis, effect size, error probability, and degrees of freedom. In the perspective of BD in an applied field, as aviation security, it seems reasonable to have a focus on the reduction of the probability of Type II error (or β -error, false negative). This, however, does not imply that Type-I error (or α -error, false positive) is of less importance. But when it comes to preventing an act of terrorism it is essential not to miss any deceptive behaviour (e.g., false negative). Reducing the false negative is possible by enhancing power. This can be achieved by increasing the sample size to an appropriate level (Price et al., 2015). Aiming for the same level of power for different fields of research is not recommendable (Baguley, 2004). Referring to practical applications, for instance in aviation security, justice or preventing criminality, it seems advisable to increase the power (e.g., up to 0.9) in order to decrease Type II error (Kyonka, 2019). Next to this, the effect size which indicates the magnitude of the observed statistical relationship (e.g., Price et al., 2015) and the actual effect which is related to practical implications (Baguley, 2004) are relevant. Since operational applications, as BD, need to be well justified and scientifically understood, it seems reasonable, especially for field studies that should result in implementing or enhancing security procedures, to aim for rather a medium effect size.

Due to the fact that future results, for example indicators of deceptive behaviour, will be applied and utilized by human operative personnel in the field, small effects on behavioural differences, which are not detectable to the human eye, are of less interest and benefit in regard to BD. Particularly very small effects might be less effective, when the human operator solely detects and recognizes, such as indicators of deception, movement patterns, and further nonverbal behavioural cues via CCTV, without intelligent technical assistance. Vice versa, the same is true for an overpowered study design that might indicate significantly many statistical differences with small effects which has no practical implication or further benefit (Kyonka, 2019). This means that a valid, sufficiently large sample size is of major importance, even to accomplish medium effect sizes. Consequently, the calculation of the sample size is required before conducting the study in all cases, so precise planning is recommended (e.g., Baguley, 2004; Bakker et al., 2016; Cohen, 1990). For a detailed discussion on effect size, see for instance Hedges (2008) or Kelley & Preacher (2012).

Including a control group in the study design allows the researcher to conduct statistical group comparisons, for example, by analysis of variance (ANOVA) and post hoc pairwise comparisons. However, research targeting BD often leads to data that does not fulfil the requirements to conduct parametric analysis. Although parametric methods are known as being relatively robust against violations, it is not true if combined violations are present, e.g., not normally distributed and unequal samples size. Given a huge sample size, it might be possible that the data belongs naturally to a normal distribution. Here, known tests for distribution, such as Shapiro-Wilk, might be too sensitive. Therefore, via plots and log transformation possible normal distribution can be revealed.

Further, when performing classical *t*-tests it is important to know that this test might be robust regarding violations of its requirements when conducted as Monte Carlo simulation. Nonetheless, it is important to bear in mind that nonparametric methods are probably the appropriate way for data analysis, although this might result in smaller effect sizes (Field, 2009).

The statistical procedures should be well selected in referring to the study design: applied methods, level of measurements, and sample size. Bearing in mind that research in the field of deception justifies on the first sight mainly qualitative approaches, most statistical procedures demand as high a level of measurements as validity data can provide. Once again, due to that, it seems recommendable to check whether the outlined project design might include some quantitative data that possibly accomplishes the need for complex analysis.

For the sake of enhancing quality of research in this particular area of interest, full data description, exact publication of statistical procedures, and results are of very high importance for the community. Transparency and well-designed statistical analysis will lead to better understanding, professional exchange, and possibly enhance operative security measures.

Review Process/Quality Check

In general, quantitative psychological studies are judged based on the calculation of the statistical quality criteria validity, reliability, and objectivity (e.g., Howitt & Cramer, 2011). Furthermore, to evaluate the scientific quality of research the factors relevance (including theoretical and practical relevance), methodological rigor (including construct validity, internal validity, external validity, and statistical validity), ethical acceptability, and reporting standards are used (e.g., Andrade, 2018; Howitt & Cramer, 2011). Furthermore, the standard for publication of scientific research is a peer-review process. In the best-case scenario, scientific research should try to fulfil these requirements. However, applied research often needs to compromise regarding scientific standards, as the results might be restricted, publishing of the results is not always possible. Nonetheless, a clearly defined protocol for the assessment of the conducted research should be followed in order to provide a consolidated estimation of the general informative value of the research.

Conclusion

This paper is intended to highlight the important aspects for conducting research in the field of behaviour analysis and deception detection in order to meet scientific requirements within the civil aviation environment. Furthermore, it is a guideline for practitioners to evaluate programs, tools, and strategies they might be offered to implement. Throughout this document, all questions considered essential in the investigation have been covered. Among others, practitioners should take into account the importance of knowing the theoretical framework of the topic to be investigated, the formulation of good research questions and hypotheses that can answer those questions, the selection and size of the sample, the operationalization of the variables to be measured, as well as the selection of the appropriate statistical analysis. It is also important to note that for all types of investigation that are designed to gain a deeper understanding on the topic of "deception detection", the full approval of each participant as well as compliance of ethical or staff council requirements have to be ensured. Neglecting any of the issues detailed above can make the results present a lack of validity, while performing according to common standards will enhance quality of research.

Conflict of Interest

The authors of this article declare no conflict of interest.

Authors' Contribution

The authors equally contributed to the elaboration of this manuscript.

References

- Andrade, C. (2018). Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian Journal of Psychological Medicine*, 40(5), 498-499. <https://doi.org/10.4103/IJPSYM.IJPSYM>
- Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics*, 35(2), 73-80. <https://doi.org/10.1016/j.apergo.2004.01.002>
- Bakker, M., Hartgerink, C. H. J., Wicherts, J. M., & van der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, 27(8), 1069-1077. <https://doi.org/10.1177/09567976166647519>
- Baumgartner, T., Gianotti, L. R. R., & Knoch, D. (2013). Who is honest and why: Baseline activation in anterior insula predicts inter-individual differences in deceptive behavior. *Biological Psychology*, 94(1), 192-197. <https://doi.org/10.1016/j.biopsycho.2013.05.018>
- Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: An overview. *Frontiers in Psychology*, 3(Sep), 1-11. <https://doi.org/10.3389/fpsyg.2012.00342>
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37, 245-257. <https://doi.org/10.1037/0003-066X.37.3.245>
- Blair, J. P., & Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science & Management*, 6(2), 77-83. <https://doi.org/10.1350/ijps.6.2.77.34465>
- Blandón-Gitlin, I., Fenn, E., Masip, J., & Yoo, A. H. (2014). Cognitive-load approaches to detect deception: Searching for cognitive mechanisms. *Trends in Cognitive Sciences*, 18(9), 441-444. <https://doi.org/10.1016/j.tics.2014.05.004>
- Bogaard, G., Meijer, E. H., Vrij, A., & Nahari, G. (2022). Detecting deception using comparable truth baselines. *Psychology, Crime and Law*, 29(6), 567-583. <https://doi.org/10.1080/1068316X.2022.2030334>
- Bogaard, G., Nußbaum, M., Schlaudt, L. S., Meijer, E. H., Nahari, G., & Vrij, A. (2022). A comparable truth baseline improves truth/lie discrimination. *Applied Cognitive Psychology*, 36(5), 1060-1071. <https://doi.org/10.1002/acp.3990>
- Bryant, E. (2008). Real lies, white lies and gray lies: Towards a typology of deception. *Kaleidoscope: A Graduate Journal of Qualitative Communication Research*, 7, 23-48.
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6(3), 203-242. <https://doi.org/10.1111/j.1468-2885.1996.tb00127.x>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365-376. <https://doi.org/10.1038/nrn3475>
- Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, 71(3), 193-204. <https://doi.org/10.1016/j.ijpsycho.2008.09.006>
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312. <https://doi.org/10.1037/10109-028>
- Costa, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 2, pp. 179-198). Sage Publications, Inc. <https://doi.org/10.4135/9781849200479.n9>
- Crothers, C., & O'Brien, T. (2020). The contexts of the christchurch terror attacks: Social science perspectives. *Kotuitui: New Zealand Journal of Social Sciences Online*, 15(2), 247-259. <https://doi.org/10.1080/1177083X.2020.1746364>
- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. Springer. https://doi.org/10.1007/978-0-387-21573-0_5
- Denault, V. (2020). Misconceptions about nonverbal cues to deception: A covert threat to the justice system? *Frontiers in Psychology*, 11, 1-4. <https://doi.org/10.3389/fpsyg.2020.573460>
- Denault, V., Plusquellec, P., Juppe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., Sporer, S. L., Rioux-Turcotte, J., Jarry, J., Walsh, D., Otgaar, H., Viziteu, A., Talwar, V., Keatley, D. A., Blandón-Gitlin, I., Townson, C., Deslauriers-Varin, N., Lilienfeld, S. O., Patterson, M. L., ... Van Koppen, P. J. (2020). The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts. *Anuario de Psicología Jurídica*, 30(1), 1-12. <https://doi.org/10.5093/apj2019a9>
- Denault, V., & Zloteanu, M. (2022). Darwin's illegitimate children: How body language experts undermine darwin's legacy. *Evolutionary Human Sciences*, 4, 1-34. <https://doi.org/10.1017/ehs.2022.50>
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Diamond, A. (2018). Executive functions. *Neuropsychology for Occupational Therapists: Cognition in Occupational Performance*, 165 -176. <https://doi.org/10.1002/9781119557036.ch10>
- Docan-Morgan, T. (Ed.) (2019). *The Palgrave handbook of deceptive communication*. Palgrave Macmillan. <https://doi.org/https://doi.org/10.1080/00029157.2011.10404358>
- Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, 58(4), 723-733. <https://doi.org/10.1287/mnsc.1110.1449>
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers*, 28(1), 1-11. <https://doi.org/10.3758/BF03203630>
- Europol. (2023). *TE-SAT, European Union terrorism situation and trend report 2023*. Europol. <https://doi.org/doi/10.2813/370206>
- Evans, J. St. B. T. (2006). Heuristic-analytic theory. *Psychonomic Bulletin & Review*, 13(3), 378-395. <https://0-link-springer-com.wam.city.ac.uk/content/pdf/10.3758%2FBF03193858.pdf>
- Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling*, 11(3), 244-252. <https://doi.org/10.1002/jip.1414>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>
- Feijoo-Fernández, M. C., Halty, L., & Sotoca-Plaza, A. (2023). Like a cat on hot bricks: The detection of anomalous behavior in airports. *Journal of Police and Criminal Psychology*, 38, 72-82. <https://doi.org/10.1007/s11896-020-09371-5>
- Fernández, M. C. F., & Halty, L. (2018). It's the intention that counts. A review on deception detection focused on intentions. *Papeles del Psicólogo*, 39(1), 51-59. <https://doi.org/10.23923/pap.psciol2018.2849>
- Field, A. (2009). *Discovering statistics using IBM SPSS Statistics*. Sage Publications. <https://doi.org/10.1234/12345678>
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72(6), 1429-1439.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31(1), 58-75. <https://doi.org/10.1080/00909880305377>
- Galizzi, M. M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: A systematic lab-field study. *Management Science*, 65(3), 976-1002. <https://doi.org/10.1287/mnsc.2017.2908>
- Gerlach, P., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1-44. <https://doi.org/10.1037/bul0000174.supp>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62(1), 451-482. <https://doi.org/10.1146/annurev-psych-120709-145346>
- Harrigan, J., Rosenthal, R., & Scherer, K. (2008). *The new handbook of methods in nonverbal behavior research*. <https://doi.org/10.1093/acprof:oso/9780198529620.001.0001>
- Hart, C. L., Lemon, R., Curtis, D. A., & Griffith, J. D. (2020). Personality traits associated with various forms of lying. *Psychological Studies*, 65, 239-246. <https://doi.org/10.1007/s12646-020-00563-x>
- Harvey, A. C., Vrij, A., Sarikas, G., Leal, S., Juppe, L., & Nahari, G. (2018). Extending the verifiability approach framework: The effect of initial questioning. *Applied Cognitive Psychology*, 32(6), 787-804. <https://doi.org/10.1002/acp.3465>
- Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be assessed reliably? A meta-analysis of criteria-based content analysis. *Psychological Assessment*, 29(6), 819-834. <https://doi.org/10.1037/pas0000426>
- Hedges, L. V. (2008). What are effect sizes and why do we need them? *Child Development Perspectives*, 2(3), 167-171. <https://doi.org/https://doi.org/10.1111/j.1750-8606.2008.00060.x>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83. <https://doi.org/https://doi.org/10.1017/S0140525X0999152>
- Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal of Forensic Science*, 39(3), 793-807. <https://doi.org/10.1520/JFS13657>
- Howitt, D., & Cramer, D. (2011). *Introduction to research methods in psychology* (Third ed.). Pearson Education Limited. <https://doi.org/10.1067/mcm.2003.64>
- Huber, O. (2014). *Das psychologische Experiment: Eine Einführung* [The psychological experiment: An introduction] (6., überar.). Hogrefe Verlag.
- Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften* [Research methods for psychology and social science] (2., überar.). Springer Berlin Heidelberg.

- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2011). *Criminal interrogation and confessions*. Jones & Bartlett Publishers.
- International Civil Aviation Organization (ICAO). (2017). *Annex 17. Safeguarding international civil aviation against acts of unlawful interference*. <https://www.icao.int/Security/SFP/Pages/Annex17.aspx>
- Jenkins, B. M. (2021). The terrorist threat to transportation targets and preventive measures. In A. P. Schmid (Ed.), *Handbook of terrorism prevention and preparedness* (pp. 790-815). ICCT Press Publication. <https://doi.org/10.19165/2020.6.01>
- Jenkins, M. B., & Butterworth, B. R. (2018). *Does "See something, say something" work?* Mineta Transportation Institute Publications. <https://transweb.sjsu.edu/research/SP1118-See-Something-Say-Something>
- Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: The dark triad traits and deception. *Personality and Individual Differences*, 70, 117-119. <https://doi.org/10.1016/j.paid.2014.06.038>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, 22(2), 420-432. <https://doi.org/10.1037/a0019265>
- Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, 7(4), 409-411. [https://doi.org/10.1016/0010-0277\(79\)90024-6](https://doi.org/10.1016/0010-0277(79)90024-6)
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137-152. <https://doi.org/10.1037/a0028086>
- Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, 54(1), 1-20. <https://doi.org/10.1348/000711001159357>
- Khalidi, K. (2017). Quantitative, qualitative or mixed research: Which research paradigm to use? *Journal of Educational and Social Research*, 7(2), 15-24. <https://doi.org/10.5901/jesr.2017.v7n2p15>
- Koller, C. I., Wetter, O. E., & Hofer, F. (2015a). What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues. *Perception*, 44(6), 679-708. <https://doi.org/10.1177/0301006615594271>
- Koller, C. I., Wetter, O. E., & Hofer, F. (2015b). "Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions. *Applied Cognitive Psychology*, 30(2), 178-187. <https://doi.org/10.1002/acp.3175>
- Koller, D., Hofer, F., Grolig, T., Ghelfi, S., & Verschuere, B. (2020). What are you hiding? Initial validation of the reaction time-based searching concealed information test. *Applied Cognitive Psychology*, 34(6), 1406-1418. <https://doi.org/10.1002/acp.3717>
- König, H. (2020). *Lüge und Täuschung in den Zeiten von Putin, Trump & Co.* [Lies and deception during the period of Putin, Trump & Co.]. Transcript Verlag.
- Kyonka, E. G. E. (2019). Tutorial: Small-N power analysis. *Perspectives on Behavior Science*, 42(1), 133-152. <https://doi.org/10.1007/s40614-018-0167-4>
- Levine, T. R. (2017). Ecological validity and deception detection research design. *Communication Methods and Measures*, 12(1), 45-54. <https://doi.org/10.1080/19312458.2017.1411471>
- Li, Z., Li, X., Dong, C., Guo, F., Zhang, F., & Zhang, Q. (2021). Quantitative analysis of global terrorist attacks based on the global terrorism database. *Sustainability (Switzerland)*, 13(14), 1-19. <https://doi.org/10.3390/su13147598>
- Lipowski, E. E. (2008). Developing great research questions. *American Journal of Health-System Pharmacy*, 65(17), 1667-1670. <https://doi.org/10.2146/ajhp070276>
- Lowhorn, G. L. (2007). *Qualitative and quantitative research: How to choose the best design*. Academic Business World International Conference. Nashville, Tennessee. <https://ssrn.com>
- Lum, C., Kennedy, L. W., & Sherley, A. (2006). Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. *Journal of Experimental Criminology*, 2(4), 489-516. <https://doi.org/10.1007/s11292-006-9020-y>
- Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. *Journal of Experimental Social Psychology*, 66, 100-106. <https://doi.org/10.1016/j.jesp.2015.09.018>
- Mann, S., Vrij, A., Leal, S., Granhag, P. A., Warmelink, L., & Forrester, D. (2012). Windows to the soul? Deliberate eye contact as a cue to deceit. *Journal of Nonverbal Behavior*, 36(3), 205-215. <https://doi.org/10.1007/s10919-012-0132-y>
- McDermott, R. (2011). Internal and external validity. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Handbook of experimental and political science*. (pp. 27-40). Cambridge University Press.
- Meterko, V., & Cooper, G. (2022). Cognitive biases in criminal case evaluation: A review of the research. *Journal of Police and Criminal Psychology*, 37(1), 101-122. <https://doi.org/10.1007/s11896-020-09425-8>
- Muñoz García, A., Gil-Gómez de Liaño, B., & Pascual-Ezama, D. (2023). The role of cognition in dishonest behavior. *Brain Sciences*, 13(3), Article 394. <https://doi.org/10.3390/brainsci13030394>
- Nahari, G., & Nisin, Z. (2019). Digging further into the speech of liars: Future research prospects in verbal lie detection. *Frontiers in Psychiatry*, 10, Article 56. <https://doi.org/10.3389/fpsy.2019.00056>
- Neal, T. M. S., Lienert, P., Denne, E., & Singh, J. P. (2022). A general model of cognitive bias in human judgment and systematic review specific to forensic mental health. *Law and Human Behavior*, 46(2), 99-120. <https://doi.org/10.1037/lhb0000482.supp>
- Petticrew, M., Egan, M., Thomson, H., Hamilton, V., Kunkler, R., & Roberts, H. (2008). Publication bias in qualitative research: What becomes of qualitative research presented at conferences? *Journal of Epidemiology and Community Health*, 62(6), 552-554. <https://doi.org/10.1136/jech.2006.059394>
- Pierce, L., & Balasubramanian, P. (2015). Behavioral field evidence on psychological and social factors in dishonesty and misconduct. *Current Opinion in Psychology*, 6, 70-76. <https://doi.org/10.1016/j.copsyc.2015.04.002>
- Piquero, A. R. (Ed.). (2016). *The handbook of criminological theory* (Issue 1). Wiley-Blackwell, A John Wiley & Sons, Ltd., Publication.
- Porter, C. N., Taylor, R., & Salvaneli, G. (2021). A critical analysis of the model statement literature: Should this tool be used in practice? *Journal of Investigative Psychology and Offender Profiling*, 18(1), 35-55. <https://doi.org/10.1002/jip.1563>
- Price, P. C., Jhangiani, R. S., & Chiang, I.-C. A. (2015). *Research methods in psychology* (2nd Canadian edition). BCCampus. https://doi.org/10.5005/jp/books/12408_3
- Roeser, K., McGregor, V. E., Stegmaier, S., Mathew, J., Kübler, A., & Meule, A. (2016). The dark triad of personality and unethical behavior at different times of day. *Personality and Individual Differences*, 88, 73-77. <https://doi.org/10.1016/j.paid.2015.09.002>
- Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist*, 46(4), 409-415. <https://doi.org/10.1037/0003-066X.46.4.409>
- Serota, K. B., & Levine, T. R. (2014). A few prolific liars: Variation in the prevalence of lying. *Journal of Language and Social Psychology*, 34(2), 138-157. <https://doi.org/10.1177/0261927X14528804>
- Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, 36(1), 2-25. <https://doi.org/10.1111/j.1468-2958.2009.01366.x>
- Serota, K. B., Levine, T. R., & Docan-Morgan, T. (2022). Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both? *Communication Monographs*, 89(3), 307-331. <https://doi.org/10.1080/03637751.2021.1985153>
- Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2013). Expect the unexpected? Variations in question type elicit cues to deception in joint interviewer contexts. *Applied Cognitive Psychology*, 27(3), 336-343. <https://doi.org/10.1002/acp.2911>
- Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences*, 12(2), 48-53. <https://doi.org/10.1016/j.tics.2007.11.008>
- Stuss, D. T. (1992). Biological and psychological development of executive functions. *Brain and Cognition*, 20(1), 8-23. [https://doi.org/https://doi.org/10.1016/0278-2626\(92\)90059-U](https://doi.org/https://doi.org/10.1016/0278-2626(92)90059-U)
- Swiss Federal Intelligence Service. (2023, June 26). *Switzerland's security 2023* (press release). <https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-95984.html>
- Synergology, the Official Website. (n.d.a). *Décrypter le langage corporel avec la discipline synergologie* [Decipher body language with the discipline of synergology]. <http://non-verbal.synergologie.org/nonverbal/synergologie>
- Szymankiewicz, Ł. (2022). Evolution of aviation terrorism – El Al Israeli airlines, case study. *Journal of Strategic Security*, 15(1), 106-125. <https://doi.org/10.5038/1944-0472.15.1.1945>
- Taylor, S., & Asmundson, G. J. G. (2008). Internal and external validity in clinical research. In D. McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 23-34). Sage publications.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1021/ac60205a032>
- UK National Protective Security Authority. (2023). *Behavioural detection - best practice, guidance and advice*. <https://www.npsa.gov.uk/behavioural-detection-0>
- US Government Accountability Office. (2010). *Aviation security: Efforts to validate TSA's passenger screening behavior detection program underway, but opportunities exist to strengthen validation and address operational challenges*. GAO-10-763 (May). <https://www.gao.gov/products/GAO-10-763>
- US Government Accountability Office. (2012a). *Aviation security - 9/11 anniversary observations on TSA's progress and challenges in strengthening aviation security*. GAO-12-1024T (September). <https://www.gao.gov/assets/gao-12-1024t.pdf>
- US Government Accountability Office. (2012b). *TSA: progress and challenges faced in strengthening three key security programs*. GAO-12-541T (March). <https://www.gao.gov/products/gao-12-541t>
- US Government Accountability Office. (2017). *Aviation security: TSA does not have valid evidence supporting most of the revised behavioral indicators used in its behavior detection activities*. GAO-17-608 (July). <https://www.gao.gov/products/GAO-17-608R>

- Verigin, B. L., Meijer, E. H., Bogaard, G., & Vrij, A. (2019). Lie prevalence, lie characteristics and strategies of self-reported good liars. *PLoS ONE*, *14*(12), 1-16. <https://doi.org/10.1371/journal.pone.0225566>
- Vernham, Z., Vrij, A., Mann, S., Leal, S., & Hillman, J. (2014). Collective interviewing: Eliciting cues to deceit using a turn-taking approach. *Psychology, Public Policy, and Law*, *20*(3), 309-324. <https://doi.org/10.1037/law0000015>
- Verschuere, B., & Meijer, E. H. (2014). What's on your mind? Recent advances in memory detection using the concealed information test. *European Psychologist*, *19*(3), 162-171. <https://doi.org/10.1027/1016-9040/a000194>
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley.
- Vrij, A. (2016). Baseline as a lie detection method. *Applied Cognitive Psychology*, *30*(6), 1112-1119. <https://doi.org/10.1002/acp.3288>
- Vrij, A., Fisher, R. P., & Leal, S. (2023). How researchers can make verbal lie detection more attractive for practitioners. *Psychiatry, Psychology and Law*, *30*(3), 383-396. <https://doi.org/10.1080/13218719.2022.2035842>
- Vrij, A., Granhag, P. A., Ashkenazi, T., Ganis, G., Leal, S., & Fisher, R. P. (2022). Verbal lie detection: Its past, present, and future. *Brain Sciences*, *12*(12), 1-20. <https://doi.org/10.3390/brainsci12121644>
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253-265. <https://doi.org/10.1007/s10979-007-9103-y>
- Walczyk, J. J., Tcholakian, T., Newman, D. N., & Duck, T. (2016). Impromptu decisions to deceive. *Applied Cognitive Psychology*, *30*(6), 934-945. <https://doi.org/10.1002/acp.3282>
- Weinberger, S. (2010). Intent to deceive ? *Nature*, *465*(7297), 412-415.
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*, Article 1832. <https://doi.org/10.3389/fpsyg.2016.01832>
- World Medical Association declaration of Helsinki. (2013). Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191-2194. <https://doi.org/10.1001/jama.2013.281053>
- Yadav, D. K., & Nikraz, H. (2014). Implications of evolving civil aviation safety regulations on the safety outcomes of air transport industry and airports. *Aviation*, *18*(2), 94-103. <https://doi.org/10.3846/16487788.2014.926641>