VOLUMEN 26
SEPTEMBRER - DECEMBER
2005
92

Sumario
Contents

JOURNAL OF THE GENERAL COUNCIL OF PSYCHOLOGICAL ASSOCATIONS

Special Section

Sección monográfica

PAPELES DEL PSICOLOGO

Publisher
Colegio Oficial de Psicólogos/Spanish Psychological Association, Spain

Editorial Committee
Junta de Gobierno del Colegio Oficial de Psicólogos/Policy Committee of the Spanish Psychological Association

Dean
Francisco J. Santolaya Ochando

Vice-Dean 1
Eduardo Montes Velasco

Vice-Dean 2
Manuel Berdullas Temes

Secretary
Manuel Mariano Vera Martínez

Assistant Secretary
Julián Baltasar Jaume

Treasurer
Juan Manuel Egurtza Muniain

Board members
Rosario Carcas, Fernando Chacón, Aurora Gil Álvarez, José Ramón Fernández Hermida, Lorenzo Gil, Rosa Jiménez Tornero, Margarita Laviana, Cristina López Díaz, Manuel Enrique Medina, Manuel Rodríguez Fernández y Concepción Santo Tomás de Abajo.

Advisory Committee
José Antonio Aldaz, Esteban Alonso, Isaac Amigo, José Arévalo Serrano, Pilar Arránz, José María Arredondo, Dositeo Artiaga, Mª Dolores Avia, Sabino Ayestarán, Francisco Bas, Vicent Bermejo, Amalio Blanco, Cristina Botella, Carmen Bragado, Gualberto Buela, José Buendía, Vicente Caballo, Francisco Cabello, José Cáceres, Rosa Calvo, Fernando Calvo, Amalia Cañas, Antonio Capafons, José Carlos Caracuel, Helio Carpintero, Mario Carretero, José Antonio Carrobles, Miguel Costa, Antonio Coy, Sixto Cubo, Piedad Cueto, Fernando Díaz Albo, María José Díaz-Aguado, Jesús A. De Diego, Raúl De Diego, Andrés Duarte López, Rocío Fernández Ballesteros, Nicolás Fernández Losa, Jorge Fernández Del Valle, Concepción Fernández Rodríguez, Alfredo Fornos, Enrique García Huete, Miguel Anxo García Álvarez, César Gilolmo, Jesús Gómez Amor, Jorge L. González Fernández, Julio Antonio González García, José Gutiérrez Terrazas, Adolfo Hernández Gordillo, Florencio Jiménez Burillo, Cristóbal Jiménez Jiménez, Annette T.Kreuz, Francisco Javier Labrador, José Carlos León Jarriego, Jesús Ramón Loitegui, Roberto Longhi, Aquilino Lousa, Araceli Maciá, Emiliano Martín, María Angeles Martínez Esteban, José Joaquín Mira, Luis Montoro, José Muñiz, Nicomedes Naranjo, Conrado Navalón, José Ignacio Navarro Guzmán, Luis De Nicolás, Soledad Ortega Cuenca, Pedro Pérez García, Marino Pérez Álvarez, Félix Pérez Quintana, José Luis Pinillos, José Antonio Portellano, José María Prieto, Ismael Quintanilla, Francisco Ramos, Jesús Rodríguez Marín, Carlos Rodríguez Sutil, José Ignacio Rubio, Carlos Samaniego, Aurelia Sánchez Navarro, Javier Urra, Miguel Angel Vallejo y Jaime Vila.

Editor
Serafín Lemos Giráldez

Associate Editors
José Ramón Fernández Hermida, Manuel Enrique Medina Tornero, José Carlos Núñez Pérez y José María Peiró Silla

Design and typesetting
Juan Antonio Pez Martínez

Editorial and advertising office
Juan Antonio Pez Martínez
Consejo General de Colegios Oficiales de Psicólogos
C/ Conde de Peñalver, 45-5º Izq.
28006 Madrid - España
Tels.: 91 444 90 20 - Fax: 91 309 56 15
E-mail: papeles@correo.cop.es

Printer
Intigraf S.L.
C/ Cormoranes, 14. Poligono Industrial La Estación.
28320 Pinto Madrid

Legal registration code
M-27453-1981 / ISSN 0214-7823

46,600 copies of Issue 92 of Papeles del Psicólogo were produced.
The publishers are not responsible for the opinions expressed in the articles reproduced in the journal.

Papeles del Psicólogo is included in Psicodoc and in the ISOC (Psedisoc) and IBECS databases.
It can also be viewed on the website of the Spanish Psychological Association (*Colegio de Psicólogos*): http://www.cop.es

Papeles del Psicólogo, 2005. Vol. 26, pp. 57-58

MALINGERING, DECEPTION AND LIES

Serafín Lemos Giráldez
Psychology Faculty, University of Oviedo

Malingering, deception and lies clearly form part of human nature, and are present in both the personal sphere and in life within society in general. From the briefest of structured greetings, of the type "Good morning, how are you? Fine, thanks", where words lose their genuine meaning in mere polite formulas, to the most sophisticated forms of communication in the complex social framework, we are involved in a game of roleplay, where things become mixed and confused: what we are with what we appear to be, reality with image, the function of author with that of actor.

Deception is not exclusive to the human species, but is a characteristic found in all primates and in other animals living in highly complex social environments; in the animal and plant kingdoms, numerous types of living creatures have developed, in the course of evolution, highly sophisticated capacities for camouflage and adaptation, which have prospered by virtue of the confusion created in competitors and predators.

Survival in a complex social medium has favoured the development of the cerebral neocortex in the human being and in other higher mammals, which has in turn made possible the acquisition of extraordinary mental abilities and concepts, such as self-awareness and the theory of mind, which permit not only the recognition of personal characteristics, but also the anticipation of the thoughts and intentions of congeners, thus increasing social skills and group cohesion. Without entering a discussion of whether primates are conscious of their behaviour, their thinking is reflective and their acts follow a preconceived plan (the mentalist hypothesis), or whether these abilities are the result of mere innate reflexes or instrumental learning processes (the behaviourist hypothesis), what seems evident is that these capacities are at the basis of what Whiten and Byrne (1997) have called "Machiavellian intelligence", a theory according to which the primates have managed to develop diverse social strategies that are advantageous to their survival, resorting for their own convenience to the use of agonistic or cooperative behaviours, according to the demands of the situation. Machiavellian intelligence is a capacity that appears to have been induced by the need to master ever more refined forms of manipulation and fraud in the social context, and which manifests itself through the use of strategies of tactical pretence, lying and deceit. As Smith (2005) argues, Machiavellian intelligence may have provided the driving force for our ancestors to acquire ever greater intelligence and to increase our tendency to change our minds, to make deals, to boast/bluff and to plot with others. Smith thus considers human beings to be born liars, having developed much more sophisticated forms of deception than even our closest primate relatives.

But pretence and the deceiving of others would not have reached such a degree if we humans had not also developed the ability to deceive ourselves. Self-deception helps us to lie to others more convincingly, and the capacity for believing our own lies helps us to more effectively dupe those around us. Furthermore, it permits us to perfect the art of "lying sincerely", without the need to resort to theatricality to pretend that we are telling the truth. This is the thesis of sociobiologist Robert Trivers (2002), who argues that the chief function of self-deception is to be able to deceive others more easily, so that credulity with regard to one's own fabrication makes it more convincing for everyone else.

Thus, pretence, implicit lying and deliberate deceit form part of all the scenarios in which human social life unfolds. In an ongoing developmental process that begins in childhood, we lose spontaneity as we gradually become convinced that honesty is not always possible or appropriate, because it can harm other people or oneself. So, well-intentioned friends lie in order to flatter, to sweeten the truth, to give support or to protect; politicians and social leaders lie to achieve their aims, to avoid problems or to seduce the electorate (with the paradox that it is the biggest liars who are keenest to expose the lies of their adversaries); the media lie, concealing information or publishing information that serves their interests, emphasizing certain news items or counteracting them with others; publicists and salespersons lie in all types of commercial transaction in order to win over their clients; and, along with many others, professionals lie so as to defend their interests, or to achieve social recognition or the satisfaction of their clients. In sum, everyone tries to accommodate reality to their own intentions, expectations or needs; but what is most sur-

prising is that knowing the world is like that, we act as though everything was true – or perhaps we need to persuade ourselves that it is.

In the different contexts of the psychologist's professional activity, dissemblance, concealment, exaggeration, leaking and falsification of the information provided are highly frequent phenomena, and constitute important obstacles to the proper assessment of cases and the decision-making process. Undoubtedly, the same difficulties exist in many other professional fields; in our own context, however, such behaviours can have a range of causes, which may be pathological (the existence of a mental disorder), criminological (the intention to avoid legal responsibility) or merely adaptive (the desire to achieve particular objectives in adverse circumstances) (Rogers, 1997).

An essential premise for the professional exercise of the psychologist is the cooperation and honesty of the client or patient; correct psychological assessment and diagnosis depend on the assessed person's honesty and will to offer information, as well as on the accuracy and veracity of the data provided. Although in professional practice psychologists tend to assume the truth of the testimony and data given by clients or patients on describing their behaviours, their states, their symptoms or their psychological problems, this assumption may be somewhat naive. As is well known, numerous factors may give rise to reluctance and to lack of cooperation, such as the pursuit of a particular aim (economic, professional or judicial), doubts about the confidentiality of the data, disagreement with the point of view or values of the professional, the defence of one's own interests, the nature of the assessment or test (voluntary or imposed), or simply lack of attention in performing tests or filling out scales and questionnaires. Malingering and defensive or deceitful attitudes are not, however, dichotomic phenomena, but rather tend to present varying degrees of intensity, depending on the circumstances or motives behind them.

With the aim of analyzing these issues, we have invited experts from different fields of professional psychological activity to express their points of view on how malingering, deception and lies can affect the validity of psychological assessment, and to identify the procedures and strategies employed for counteracting their effects.

The articles by Ramón Arce and Francisca Fariña and by Verónica Godoy and Lorenzo Higueras deal with a highly controversial issue, that of the credibility of testimony in forensic contexts. The key question concerns whether the credibility of a statement can be the object of scientific research, or whether it belongs to the realm of subjectivity. Arce and Fariña present a systematized procedure developed by the authors themselves for making decisions about the reliability and validity of declarations, or the veracity of the psychological trace adduced or refuted by claimants; Godoy and Higueras, on the other hand, undertake a critical examination of the validity of a procedure for determining the credibility of statements, namely, Criteria-Based Content Analysis (CBCA), employed by some forensic psychologists. The article by Jaume Masip deals exhaustively with another important question, of potential relevance to the forensic context, namely, the reliability of the popular belief that lying can be better detected through non-verbal behaviour than through the analysis of verbal messages.

Another two articles examine malingering and deception in the clinical context. The work by Mercedes Inda and cols. offers some conceptual and methodological reflections relevant to the study of malingering behaviours, as well as describing the clinical conditions in which such behaviours most frequently occur and presenting some instruments specifically designed for their exposure. Manuel Porcel and Rubén González, openly assuming that lying and pretence constitute an essential part of human behaviour, argue that pathological behaviours are nothing but a fictional cover for life problems, and that psychotherapeutic intervention, if it is to be successful, should operate on the basis of this assumption.

The analysis and management of malingering and of social desirability in psychological assessments carried out in organizational contexts are splendidly dealt with by Jesús F. Salgado, while finally, deception and acquiescence or social desirability responses on the MMPI-2 are examined in the paper by Héctor González Ordí and Iciar Iruarrizaga.

We thank all the authors for their participation in this special issue, and hope that the contributions presented here will be of interest and practical utility for professional psychologists.

REFERENCES

Rogers, R. (Ed.). (1997). Clinical assessment of malingering and deception. New York: Guilford Press.

Smith, D. L. (2005). Mentirosos natos. Mente y Cerebro, 14, 64-69.

Trivers, R. (2002). Natural selection and social theory: Selected Papers of Robert Trivers (Evolution and Cognition Series). New York: Oxford University Press.

Whiten, A. & Byrne, R. W. (Eds.). (1997). Machiavellian intelligence II: Extensions and evaluations. Cambridge: Cambridge University Press.

Papeles del Psicólogo, 2005. Vol. 26, pp. 59-77

PSYCHOLOGICAL EVIDENCE IN COURT ON STATEMENT CREDIBILITY, PSYCHOLOGICAL INJURY AND MALINGERING: THE GLOBAL EVALUATION SYSTEM (GES)

Ramón Arce* and Francisca Fariña**

** Department of Social Psychology, University of Santiago de Compostela*

***Basic Psychology Department, University of Vigo*

This paper presents a protocol adapted to the Spanish judicial system for the assessment of malingering and deception in testimony, the Global Evaluation System (GES). The GES is the product of a combination of different strategies of assessment of the reliability (CBCA and Reality Monitoring) and validity (SRA and SVA) of statements, as well as the assessment of psychological injury and malingering. Moreover, this procedure has been validated for different cases and types of witness (children, adults and the disabled), and we review the productive and effective interview protocols for the assessment of credibility and psychological injury.

El presente trabajo presenta un procedimiento de detección de la mentira y el engaño en procesos judiciales adaptado al sistema judicial español: el Sistema de Evaluación Global. El S.E.G. resulta de un compendio de diversas estrategias de análisis de la fiabilidad (CBCA y Reality Monitoring) y validez de la declaración (SVA y SRA) a la vez que de la huella psicológica y de la simulación de la misma. Además, para poder obtener unos protocolos susceptibles del presente análisis se revisan y presentan los diversos modos productivos y efectivos en la obtención de la declaración con adultos, menores, discapacitados así como la entrevista clínico-forense para la valoración del daño psíquico y de la simulación.

The formation of legal judgements and their subsequent materialization in judicial sentences, which constitute the cornerstone of the Judicial System (Sallmann and Willis, 1984), rest on the two basic dimensions formulated by the *Information Integration Models*: reliability and validity (Ostrom, Werner & Saks, 1978). Briefly, a judgement is an assessment of the evidence in one dimension. Judgements are based on a set of beliefs about the evidence (e.g., inferences about the accused's motives, or capacities) that are relevant to the assessment dimension, so that each belief has a weight that affects assessment of the evidence for the judgement dimension. This weight is known as the scalar value of the belief. But not all beliefs contribute in equal measure to the assessment of the evidence: this contribution derives from the estimation of the reliability and validity of the belief. Reliability in the courtroom is basically defined by the credibility of the witnesses; validity is determined by the relevance of the evidence to the judgement to be made. However, credibility of witnesses contributes the highest scalar value in the verdict reached by both juries (Arce, Fariña & Real, 2000) and judges (Arce, Fariña, Novo & Seijo, 2001), and is the key to reaching conflicting verdicts (e.g., attribution of high credibility to a victim's testimony is a highly robust predictor of guilt, while lack of credibility for a victim's testimony reliably predicts a not-guilty verdict).

Estimation of the credibility of a testimony would be the appreciation of the accuracy or correctness inspired in the assessor by the witness or a part of his/her statement, which induces the assessor to believe that events occurred as stated (Mira, 1989). Two models have been formulated for the estimation of credibility: the subjective or social model and the objective or scientific model (Vrij, 2000). The social model is understood as the recourse to subjective indicators –not scientifically validated– of credibility, while the objective model rests on the assumption of empirical criteria of credibility. In everyday judicial practice, both experts, that is, judges and magistrates (Piñeiro, 2005), and lay persons in law (juries) (Arce, Fariña & Seijo, 2003) use subjective criteria for estimating the credibility of testimonies, which are ineffective for the correct classification of their accuracy. Therefore, courts find it necessary to employ qualified agents for assessing the credibility of witnesses in an empirical manner by means of productive and scientifically validated techniques. In relation to

Correspondence: Ramón Arce. *Facultad de Psicoloxía. Departamento de Psicoloxía Social, Básica e Metodoloxía. Universidade de Santiago de Compostela. 15706 Santiago de Compostela. Spain. E-mail: psarce@usc.es*

this, a review of the literature revealed that the procedures carried out by psychologists based on content analysis of statements were the most effective, classifying correctly, in simulation contexts (that is, with simulated witnesses in the laboratory), between 65% and 85% of statements, while in field studies (i.e., in real conditions) effectiveness reached 100% (Vrij, 2000). Nevertheless, these instruments present some limitations that should be addressed. First, the categories of the diverse procedures proposed for the analysis of content are not homogeneous across systems. Second, they do not usually cover all types of witness (they tend to be proposed for children). Third, these procedures are not accompanied by protocols suitable for assessing evidence in each context (minors, adults, the disabled). Fourth, these procedures do not analyze the entire judicial task (they tend to focus on the alleged victim, ignoring the alleged aggressor, who may be the true victim). Fifth, they are not oriented to the search for psychological injury resulting from the crime: post-traumatic stress disorder (PTSD). Sixth, they do not include appropriate methodologies for clinical assessment in the forensic context in which malingering, or clinical lying, has to be ruled out. And seventh, systems based on statement content analysis proposed for the study of credibility are semi-objective techniques that need adjustment for achieving greater objectivity.

On the basis of the literature on the empirical assessment of statement credibility, and with the aim of addressing the limitations referred to above, we have created (Arce & Fariña, 2002, 2005, in press) a forensic psychological protocol, the result of extensive research and professional experience, which attempts to deal comprehensively with the task, the *Global Evaluation System*. In its description we shall begin with a brief review of the protocols for obtaining a statement, which constitute essential requirements for applying the categorial systems of content analysis. We shall then present the most productive and effective systems for the analysis of credibility based on content analysis. Thirdly, we shall describe a protocol for assessment of psychological injury resulting from criminal acts and for the detection of malingering. Finally, we shall combine all of the above in a procedure adapted to the Spanish legal context for assessing the reality of memory trace and psychological injury with control of malingering: the *Global Evaluation System*.

OBTAINING THE STATEMENT

The basic tools for obtaining information from witnesses are interrogations and interviews. It is well known that the success of the interview or interrogation will depend on factors such as interviewer expertise, interviewee's degree of cooperation, time elapsed since the event, and obviously, type of interview (see Memon & Bull, 1999). Interrogations constitute the instrument par excellence for obtaining statements in police and judicial inquiries, but the testimony obtained through them is not productive for the application of systems for the analysis of credibility based on the content of the statements. Likewise, not all types of interview are valid for these purposes. Indeed, structured or semi-structured interviews may introduce misleading information in the accounts of truthful persons (e.g., Loftus, Korf & Schooler, 1988), so that the distortions would derive not from lying, but from the information introduced by the interviewer. Consequently, what is required are interviews in which the interviewer does not influence the information to be recalled by the witness. Interviews in free narrative format, such as cognitive interviews, fulfil this criterion. With regard to these, Köhnken, Milne, Memon and Bull (1999), after a review of the literature, observed that they facilitated higher rates of retrieval of information, especially correct (36%), but also incorrect information (17.5%). However, this does not mean that overall accuracy is greater in one type of interview than in another. In fact, average percentage of correct information for police interviews (interrogations) is 82%, while in the cognitive interview it is 84%. In this line, researchers in the field have proposed narrative interview formats adapted to the context and to the interviewee's capacities. We shall now briefly review each one of these formats.

COGNITIVE INTERVIEW[1]

The cognitive interview comprises four general techniques for retrieval of memories:

a) The first of these techniques consists in mentally reconstructing the physical and personal contexts existing at the time of the crime (or the event to be recalled), that is, the *reconstruction of contexts*. This involves asking the witness to situate him/herself mentally at the scene of the event, taking into account:

[1] Interviews must be recorded for subsequent content analysis

-Emotional elements (e.g., try to recall how you felt).
- Sequential elements (e.g., think what you were doing at the time).
- Perceptual characteristics (e.g., think of the scene of the crime and draw the room. What did it smell of? What did you hear?).

The reasoning behind this first technique is Tulving's principle of specific encoding, that is, the contextual information of an event is encoded together with the event and connected in an associative way (Tulving & Thompson, 1973). In turn, verbal recall of the event depends on the degree to which the contextual cues of the situation to be recalled overlap with properties previously encoded (Tulving, 1983). This first cognitive interview technique is similar to that used by judges and the police, the so-called reconstruction of the events, except that the reconstruction in the cognitive interview is carried out mentally.

b) The second technique, *free recall*, consists in asking the witness to recount everything that happened, absolutely everything, including partial information and seemingly trivial or insignificant details, since these can lead to others, associated in the memory, which are indeed relevant. This strategy is especially important when it is necessary to combine the information from different witnesses. Moreover, small details, in certain cases, can produce substantial clues.

By means of these first two techniques, retrieval of the mental image and reporting of all accessible details, an initial version of events is obtained. This statement, therefore, is of a narrative type, leaving the witness to speak without interruptions or questions. It is important to point out the need to ensure, throughout the interview, an appropriate environment for witnesses to concentrate, without noise or people to distract them, and the interviewer must obviously gain witnesses' confidence so that their testimony is as truthful and productive as possible.

c) The third technique, *change of perspective*, involves trying to encourage witnesses to put themselves in the position of the victim, or of another witness to the event –even the suspect–, and to report what they would have seen if they were in the position of that other person. This technique is based on the work of Bower (1967), who found that subjects, on imagining themselves as characters in a story, recalled more details pertaining to the perspective of the character with whom they had identified than those pertaining to other characters. In this way a second version of the interview is obtained, from a different perspective.

d) The final component is the instruction that invites recall from different starting points, *reverse-order recall*. In other words, the subject is asked to recount the event in a different order (e.g., from the end to the beginning, from the middle, backwards), with the aim of recovering small details that might have been lost in a narration of events simply in the sequence that they occurred. It is attempted through this technique to reduce the effects of previous knowledge, expectations and schemata on recall, and it may also be effective for eliciting additional details (Memon, Cronin, Eaves & Bull, 1993). In support of the use of this technique Bower and Morrow (1990) point out that we tend to recall the schema or mental model we form of an event, more than the event itself.

Application of the cognitive interview is not restricted to the reproduction of a single event, but can be extended to for recall of events that occur frequently in a similar manner (Mantwill, Köhnken & Ascherman, 1995). The cognitive interview includes, in turn, supplementary techniques such as:

a) Memory gymnastics for recall of physical appearance: Did the intruder remind of anyone you know? Was there anything unusual about his/her appearance?
b) Names: Try to remember the first letter of the name, going through the alphabet one letter at a time.
c) Objects: Describe the objects that were inside and outside the room. Did they look as if they were heavy?
d) Conversations and speech characteristics: Did you hear foreign or unusual words? Did you notice any accent? Did they stutter?
e) Car registration plates: Did the numbers or letters of the registration plate remind you of anything? Were they high or low numbers?

Apart from this standard version of the cognitive interview, Fisher & Geiselman (1992) proposed an improved version, adapted to the judicial context. Nevertheless, the effectiveness and procedure in cognitive terms are the same. The following phases summarize its general improved structure.

- Phase 1. Introductions and personalization of the interview (introductions, addressing interviewee by his/her name).
- Phase 2. Establishing communication (creating a pleasant atmosphere and a rapport by means of neutral questions).

- Phase 3. Explanation of the purpose of the interview.
- Phase 4. Reconstruction of contexts.
- Phase 5. Free recall.
- Phase 6. Preparation for the interrogation (interviewee is asked to concentrate hard, to say what comes to mind just as it comes, without *making it up*; to say, if necessary, *I don't understand, I don't know, I don't remember*; to activate and compare images).
- Phase 7. Interrogation compatible with the witness (each witness has a different memory sequence of the event, to which the interviewer must adapt).
- Phase 8. Recall from different perspectives.
- Phase 9. Reverse-order recall.
- Phase 10. Summary (made by the interviewer according to what the interviewee has reported).
- Phase 11. Close (emotional warming-down and dissipation of tensions in the interviewee).

INTERVIEWING CHILD WITNESSES

When witnesses are children, some important points should be borne in mind on using the cognitive interview procedure. First of all, one of the most consistent findings with regard to children's memory refers to performance. Briefly, in free recall tasks children retrieve significantly less information than adults: levels of detail and accuracy in recall of an event increase with age (Davies, Tarrant & Flin, 1989). Second, when the recall task is related to a meaningful and familiar context, the memory abilities displayed by children are greater than when the context is unfamiliar and without meaning for them (Bauer & Mandler, 1990). In other words, in these contexts the information they retrieve is not necessarily less productive and accurate. Given that the main objective of the cognitive interview is to increase the quantity of information recalled, this is the most appropriate procedure to employ with children (Memon & Bull, 1991). Nevertheless, it is necessary to modify the instructions so that the child understands what is being asked of him/her and to adapt the demands of the task to the child's mental capacities.

With the aim of adapting the cognitive interview for children, the British authorities (Home Office and The Department of Health, 1992) appointed Professors Diane Birch and Ray Bull to draw up a protocol for obtaining statements from child witnesses. On the basis of the draft prepared by these professors and discussions with technicians, a specific protocol was drawn up for working with children. Prior to the interview itself, it is recommended to collate information on the child's developmental state, level of language, and physical, social and sexual maturity. The protocol consists of four phases to be implemented by the interviewer in the following specific order: understanding and rapport, free recall, interrogation and conclusion. The first phase, *understanding and rapport* with the child, is of fundamental importance for obtaining the statement. If the child is not relaxed and comfortable, there will not be a good flow of communication. Moreover, it should be borne in mind that children are generally brought up not to speak to strangers. Therefore, it is recommended to begin by talking about *neutral* topics, such as hobbies, friends or school. Interviewers should also be especially careful with children who might *feel guilty* or think they have done *something wrong*. Thus, the interviewer must insist upon and make very clear the need for the interviewee to tell the *truth*. Finally, the object of this first phase is to inform the child about what is expected of him/her from that point on.

In the second phase, *free recall*, witnesses are asked to recount everything that happened (e.g., Is there anything you'd like to tell me? Do you know why we're here?). The interviewer can act as a facilitator, but must never formulate specific questions. Particular attention should be paid to the child's possible cognitions. In fact, child witnesses often think the adults already know what happened, or believe they should not give information about the events (as pointed out previously, children are generally told not to talk to strangers). Therefore, it is important to stress to them that they must tell everything. Throughout the interview, the interviewer should adopt a posture of active listening, resisting the temptation to intervene during pauses and long silences.

In the third phase, *interrogation*, the following order of priority should prevail in the formulation of questions: open questions, specific but not leading questions, closed questions and deep questions. Once the interviewer is totally satisfied that the free recall is finished, he/she may, where appropriate, pose open questions so that the child provides more information on some points that need clarification. However, it is important to bear in mind that questions of the form *why?* may cause feelings of guilt in the child, so that some caution should be exercised. Moreover, both reformulation of questions and requests for repetition of a response should be avoided, given that these can be interpreted as criticism, or that the answer was wrong,

respectively. The purpose of specific questions will be to clarify certain responses previously obtained. Even so, effective control should be exercised over inherent *suggestion* in this type of question –that is, the question should not imply its own answer. Likewise, questions with bipolar response alternatives (such as *yes* vs. *no*) should be excluded at this point of the interview. In any case, the content of the questions will be mediated by the child's level of development. Closed questions, on the other hand, will be employed if the previous types have failed to produce the desired results. Questions with only two response alternatives should be avoided as far as possible, given that children display a tendency to choose the first available option, especially if it is *yes*; therefore, if this type of question becomes unavoidable, witnesses should be provided with an escape route, such as *don't know* or *don't remember*. Finally, the interviewer may formulate deep questions, which are those whose reply is implicit in them. In contrast to the proposal of the original protocol, we feel it inadvisable to ask about the identity of the perpetrator of the crime, for two reasons. First of all, it is not the expert investigator's task to identify the criminal, but rather to reach a judgement on the reliability of the facts described. Second, it would be imprudent to give names, since our procedures for the assessment of reality are for events, so that the intentional transposition of persons cannot be detected by them.

The fourth phase, the *close of the interview*, consists in a recapitulation in which the interviewer inquires, using language adapted to the child's level of development, whether what has been reported in the interview is correct; and a rounding off, in which the aim is to bring down the levels of anxiety and tension (as at the beginning of the interview, the interviewer employs neutral questions, thanks the interviewee for his/her cooperation, and lets him/her know that he/she has been of help).

To this general protocol, we have added some complementary procedures for specific cases. First, when dealing with information on concepts that the child had not yet operativized adequately, such as the estimation of time or length, we resort to comparisons with well-established referents. Thus, if we want to know the duration of a particular event, we can compare it with school breaktime (e.g., was it shorter, longer or the same length as breaktime? If the answer is that it was shorter, it probably lasted less than half an hour). Second, we employ complementary types of language for those cases in which the child's level of linguistic development advises it; that is, there are times when the minor has the images of the facts in mind, but lacks the linguistic skills to transform them into words, or this lack of ability makes the statement very short, and thus unsuitable for our purposes. In these cases, using other communication procedures with the child is not only appropriate, but also advantageous. The communication systems we employ are drawings or acting out of the mental images in question. Obviously, these types of communication are complementary, and should not be used in isolation, since they can easily lead to misinterpretation. However, when child witnesses describe or try to describe events or actions they do not understand, and for which they lack vocabulary, they can draw them, point to them on a model or doll or on themselves, act them out, or represent them using objects. It should be borne in mind that it is inadvisable to use *anatomically correct dolls*, since these have a high probability of causing errors in the expert's interpretations (Dammeyer, 1998), and because the memory elicited through this type of figure also increases the number of incorrect responses among younger children (Goodman et al., 1997). In any case, it should be clear that the information obtained by these means can only be for clarification and complementary, and would never in itself constitute evidence for prosecution.

How can we decide on whether to use the cognitive interview or the children's protocol? The basic difference between the two procedures resides in the cognitive abilities demanded of the child. The cognitive interview requires, for example, the capacity for empathy, for change of perspective. Therefore, if this capacity is assumed to be acquired gradually from age 8 or 9 (Vrij & Winkel, 1996), it is advisable to use the protocol for children with minors under 8. In any case, children under 7 have difficulty following the techniques involved in the cognitive interview (Memon, Cronin, Eaves & Bull, 1996). And even if the cognitive interview may be effective for those over age 7, there are some risks, such as those deriving from the fact that the responses can be demand-led (Memon, Wark, Bull, & Köhnken, 1997). In cases of doubt, it is advisable to begin with the cognitive interview and, if a lack of cognitive skills is noted in the interviewee, to change to the protocol for children. We have found this system to be practicable, and that it does not lead to distortions.

INTERVIEWING DISABLED PERSONS

The scientific literature on interviewing disabled persons is extremely scarce, and there is a glaring need for more studies on the specific techniques to be applied (Bull, 1995). In fact, research has found that through the cognitive interview, 32% more correct information is retrieved, but that at the same time, there is a significant increase in the quantity of confabulation. In any case, of even more concern are the problems related to leading questions, to closed questions and to deep questions. In this line, Cahill et al. (1988, cited in Bull, 1995) have drawn up a list of aspects to be avoided by interviewers:

a) Acquiescence of witnesses with leading or suggestive questions, so that the reply is that which was asked for.
b) Undue pressure that leads witnesses to confabulate (e.g., to feel as though they were part of an event they did not in fact witness).
c) Repeatedly asking questions about a particular point, leading witnesses to make conjectures or deviate from their initial response (repeated questions lead interviewees to assume that their first answer was not correct).
d) Haste in labelling the language used by these witnesses as ambiguous or inadequate.
e) Offering descriptions to witnesses with difficulties for finding their own words (e.g., "if the jacket was not dark or light, then, would you say it was a kind of brown?").
f) Providing witnesses with closed response alternatives (e.g., "Did he have a revolver or a rifle?).
g) When the witness uses a tag such as *you know?*, the interviewer should proceed in a way that allows the extraction of information (one possibility is for the interviewer to tell the witness directly that he/she does *not* know, and that he/she wants the witness to explain).
h) Ignoring a previous fragment of information from the witness that does not fit with the interviewer's assumption of how events occurred.
i) Failure to understand everything the witness relates.
j) Failure to check, using all appropriate means, that the witness has been understood.

Given the current state of the literature, and in response to demands from the courts, Arce, Novo and Alfaro (2000) drew up some guidelines and a protocol for these cases. Prior to evaluating the testimony of a supposedly disabled person, it is necessary to determine their capacity as a witness. Under the Spanish judicial system, there is in fact a general obligation to testify: the LECrim (Art. 410) states that all persons resident in Spain, be they Spanish or foreign, are obliged to respond to the call to testify in judicial proceedings. At the same time, however, it establishes a series of exemptions from this obligation, among which are physical or moral incapacity (V. Art. 417, para. 3).

The procedure begins with the application of the corresponding Wechsler scales, generally the WAIS. These provide highly reliable information on the person's intellectual capacity, and are also a robust indicator of brain lesion. Specifically, in the WAIS, we are alerted to possibly relevant lesions by those results with a significant difference between the verbal and manipulative subscales (some authors use a criterion of >10 points, while others use >15) (see Wechsler, 1976, for a review) and, in the case in question here, the manipulative coefficient should be higher than the verbal one (possible organic lesion in the left hemisphere).

Likewise, interviews and the subject's antecedents will contribute important information for defining whether or not a lesion is a possibility, as well as data on the extent of incapacity. However, not all brain lesions incapacitate subjects for giving information, even in those cases in which they may be incapacitated as courtroom witnesses.

Thus, once a possible lesion or deficit has been detected, the experts set out to identify the areas –above all the cognitive ones– affected. The recommended tools are the Test Barcelona (Peña-Casanova, 1990) and the ERFC (Gil, 1999), for their reliability and because they combine in a single instrument the evaluation of the areas that a permit a testimony to be adapted in accordance with the witnesses' limitations.

Initially, the operative lesions are classified into aphasias, alexias and acalculias. Obviously, the lesions tend to relate to one of these. However, each grouping is subdivided into areas. After the global assessment, it is necessary to make an assessment of the adequacy of the witness's testimony (i.e., identification of accessible areas, deficient areas and forms of obtaining the testimony), and the conditions under which a reliable testimony can be obtained (e.g., a statement made to a forensic psychologist expert in this type of assessment may be valid, but not one given to judges, police or lawyers). Thus, for example, a person with anomia and orophonatory praxis may display difficulties in verbally identifying behaviours and may mix up words (e.g., bandage and

bondage), though this does not necessarily render their testimony unreliable.

It is also common to find witnesses who present disorders with clear implications for obtaining a statement in the *verbal memory of texts* task. Specifically, it may be that they are incapable of direct recall of the event, but still capable of responding to an interrogation on it, insofar as they lack episodic memory but not memory of the event if this is guided by questions. Consequently, they can reply perfectly well to an interrogation, and lack the capacity to create a coherent false testimony on lacking episodic memory (this condition has been observed in habitual drug users). In the most adverse of cases, when the deficit is severe, subjects may also provide useful information, since people with neurological damage, as long as their vision is intact, can store and recover visual information (Freed et al., 1989; Hart & O'Shanick, 1993; Winograd, Smith & Simon, 1982). In fact, all that is necessary is the reproduction of the context (bearing in mind that in many of them their processing is slow, so that extreme patience is required to avoid interrupting them in the middle of their search). Under this contingency, effective lying is impossible, and the crucial task is to find a system of communication that will be defined by the neurological analysis, the most accessible being reconstruction of the events. This procedure proved to be productive with different types of deficit, and has been validated for the judicial context in various cases.

THE CLINICAL-FORENSIC INTERVIEW

A final source of information with implications for the reliability of a testimony comes from the clinical context. The instruments normally used in clinical measurement are developed with a view to dealing with a *patient*. The study of malingering is not relevant, and therefore, neither structured/semi-structured interviews nor symptom lists and psychometric instruments are suitable for the purpose of checking the simulation of a mental disorder, since they actually provide information that facilitates such malingering.

For example, the question *Do you have headaches?* (first question from the SCL-90-R, Derogatis, 2002) makes it easy for the malingerer to produce a response consistent with the simulation. This type of question provides subjects with *a guiding path* for selection of the symptoms associated with a given mental disorder, so that all they need is the ability to discriminate between items belonging to one pathology or another.

The available data show that there are no references for the traditional clinical interview in which a diagnosis of simulation is reached (e.g., Rogers, 1997), while subjects are capable of effectively simulating an illness and discriminating it from others (e.g., Arce, Carballal, Fariña & Seijo, 2004; Arce, Fariña & Pampillón, 2002). Even though some psychometric instruments have scales for controlling the *validity* of the data registered, these are not sufficient for reliably establishing malingering, because: a) the malingering diagnosis is compatible with the formulation of alternative hypotheses (e.g., Graham, 1992; Roig-Fusté, 1993), b) not all malingerers are correctly classified (e.g., Bagby, Buis & Nicholson, 1995), and c) what are provided are not diagnoses, but diagnostic impressions.

Consequently, decisions based solely on this type of instrument leave the door open to the systematic commission of two types of error: false positives (classifying those who are actually ill as malingerers) and errors of omission (failing to detect as malingerers those subjects who are in fact simulating).

In view of these problems, and with the aim of minimizing the sources of error, a multi-method assessment strategy has been proposed (e.g., Rogers, 1997). This context makes room for an interview of a clinical nature that permits a diagnosis, and whose data can be compared with those obtained by other methods.

Thus, we have developed the so-called *Clinical-Forensic Interview*. This interview, carried out by a trained interviewer with psychopathological expertise, consists in asking subjects to list, in free narrative format, their current symptoms, behaviours and thoughts, compared with their state prior to the crime (i.e., GAF on axis V of the DSM-IV-TR). If the subjects do not respond of their own accord, they are asked by means of open questions, in accordance with axis V of the DSM-IV-TR (American Psychiatric Association, 2002), to talk about their family relationships, social relationships and workplace relationships, these being assessed on the appropriate scales. Through this procedure, subjects are given a task relating to knowledge of the symptoms they present, whilst for structured and semi-structured interviews, lists of symptoms and psychometric instruments they perform a task of symptom *recognition*. For this reason, the interview is not in the form of an interrogation, but rather *non-directive*, and oriented to the reconstruction of contexts. In other words, we use the procedure of open, free-narrative interview followed by reconstruction of

contexts. This kind of interview procedure showed itself to be reliable, valid and productive in the detection of simulation of post-traumatic stress disorder in cases of alleged sexual assault and harassment (Arce, Fariña & Freire, 2002), gender violence (Arce et al., 2004) and road traffic accident (Arce, Fariña, Carballal & Novo, 2006), in the detection of a non-imputable mental disorder (Arce, Fariña & Pampillón, 2002).

The clinical information obtained should be recorded and its content analyzed. The categories of analysis are the symptoms described in the DSM-IV-TR. Thus, we created a mutually exclusive, reliable and valid categorial system, of the type Weick (1985) refers to as methodical category systems. Once the registration sheets have been drawn up, the different symptoms detected are noted. While the vast majority of the symptoms, including the most adverse ones, can be reported directly by subjects, (Lewis & Saarni, 1993), some can only be observed. Consequently, two complementary methods are involved in the detection of categories: direct report from the subject and inferences made by the coders after analyzing the protocols. For example, deterioration of memory can be reported, or reflected, directly by the subject or inferred by the coder after the interview.

IN SEARCH OF THE TRUTH: CONTENT ANALYSIS OF THE STATEMENT

The review of the literature on content analysis of statements led us in the *Global Evaluation System* to consider the assessment of the credibility of witness statements according to two parameters: validity and reliability. Validity serves to establish the admissibility of the evidence for the content analysis, while reliability is related to the indicators of reality contained in the statement.

ANALYSIS OF THE VALIDITY OF THE STATEMENT

The study of the validity of the statement as evidence is made on the basis of the complete procedure (e.g., statements to the police or a judge, other testimonies, other evidence provided) and the recordings of the statements given to the expert investigators. There are two systems for the analysis of statement validity: Statement Reality Analysis (SRA) and Statement Validity Analysis (SVA).

The system known as SRA (Undeutsch, 1967, 1988), analyzes the validity of the testimony by means of the following categories:

a) Negative or Control criteria:
- Lack of internal consistency (contradictions).
- Lack of consistency with the laws of nature or science.
- Lack of external consistency (discrepancy with other incontrovertible facts).

b) Criteria derived from the sequences of statements:
- Lack of persistence (stability in time and contexts).
- Statement inconsistent with a previous statement.

On the other hand, SVA (e.g., Steller, 1989), employs the following assessment categories:

a) Psychological characteristics:
- Appropriateness of language and knowledge.
- Appropriateness of emotional expression.
- Susceptibility to suggestion.

b) Interview characteristics:
- Coercive, suggestive or leading questions.
- Global appropriateness of the interview.

c) Motivation:
- Reasons for making a statement.
- Context of the original statement.
- Pressure to present a false statement.

d) Investigation issues:
- Consistency with the laws of nature.
- Consistency with other statements.
- Consistency with other evidence.

As a criterion for global assessment of the statement, this is indicated by its best fit to one of the following categories: *credible*, *probably credible*, *indeterminate*, *probably incredible or incredible*.

RELIABILITY OF THE STATEMENT

The study of the reliability of statements –the search for criteria of reality in their content–, carried out from recordings of the statements given to expert investigators, constitutes the major contribution of Forensic Psychology to the assessment of evidence. Three categorial systems, based on content analysis, have been proposed and shown to be productive and effective for assessing the reliability of evidence: Reality Monitoring, SRA and CBCA.

Reality Monitoring, in its seminal proposal (Johnson & Raye, 1981), asserts that true statements contain more contextual (spatio-temporal) and sense-related (sounds, smells, etc.) attributes, while fabricated testimony includes more cognitive operations, i.e., idiosyncratic information (for example, I thought, I remember seeing, I felt nervous). Spörer (1997) extended the list of criteria to eight: clarity (as opposed to vagueness), perceptual information (sense-related information, such as sounds, tastes or visual details), spatial information (places, locations), temporal information (location of the event in time,

description of event sequences), affect (expression of emotions and feelings experienced during the event), reconstruction of the story (plausibility of reconstruction of the event based on the information given), realism (plausibility, realism and sense of the story) and cognitive operations (descriptions of inferences made by others during the event). The first seven of these are linked to truth, and the eighth to falsity, making this new categorization more effective. Validation of memory attributes is usually carried out by means of comparison between the results of the statement and the prescriptions of the model, but it can also be done through a process of reasoning that involves analysis of the qualitative characteristics of the memory trace, the characteristics of related traces and mnesic assumptions.

SRA (Undeutsch, 1967, 1988) also uses categories for assessing the credibility of the statement. These are as follows:

a) General, basic criteria:
- Spatio-temporal anchorage (fixing of the action in a space and time).
- Concreteness (clarity, vividness).
- Richness of detail (large quantity of details in the narration).
- Originality of the narrations (as opposed to stereotypes or clichés).
- Internal consistency (logical and psychological coherence).
- Mention of specific details of a particular type of sexual aggression.

b) Special manifestations of the above criteria:
- Reference to details that exceed witnesses' capacity (that go beyond their imagination or capacity for understanding).
- Reference to subjective experiences (feelings, emotions, thoughts, fears).
- Mention of unforeseen events or unexpected complications.
- Spontaneous corrections, specifications and additions during the statement.
- Statements that negatively affect their own interests.

With all these decision criteria a global assessment is made, in which the two factors *general criteria* and *special manifestations of the general criteria* are weighted positively towards truth, i.e., the presence of these criteria indicate that the statement is true, but their absence does not imply that it is false. For their part, the presence of the validity criteria *Control criteria* and *Criteria derived from the sequences of statement* are detrimental to the truth value of the statement. In any case, it should be borne in mind that each criterion has a limited weight in the determination of category (true vs. false) or the extent to which a statement represents a situation actually experienced by the witness. Moreover, it prescribes adherence to four maxims in the determination of whether the story relates a real event or not:
- Intensity of the comments in the different criteria.
- Number of details in the story that are related to a criterion (or more).
- Witnesses' capacity for giving evidence (age, intelligence, suggestibility).
- Characteristics of the narrative event (e.g., complexity, relevance).

In 1994, Steller and Köhnken proposed, on the basis of previous work, an integrated system of categories whose purpose was to assess the statements of minors who were the alleged victims of sexual abuse. This system, Criterion-Based Credibility Assessment (CBCA) consists of five main categories with 19 criteria for assessment:

a) General characteristics:
- Logical structure (coherence and internal consistency).
- Lack of structure (disorganized presentation).
- Number of details (abundance of different details or facts).

b) Specific content:
- Contextual machinery (situation of the narrative in space and time).
- Description of interactions (chain of actions between witness and other actors).
- Reproduction of conversations.
- Unexpected complications during the incident (e.g., unexpected interruption).

c) Peculiarities of content:
- Unusual details (details with low probability of occurrence).
- Superfluous details (irrelevant details that do not contribute significantly to the body of facts).
- Incomprehension of details accurately recounted (provision of details that the child does not understand, but that indeed make sense).
- Related external associations (inclusion of information external to the events in question, but related to them, e.g., in a case of sexual assault, recalling previous conversations about the subject).
- Reference to subjective mental state (references to

one's own feelings, emotions or cognitions).
- Attributions about the perpetrator's mental state (references to aggressor's mental state and attribution of motives).

d) Content related to motivation:
- Spontaneous corrections (spontaneous correction or improvement of the statement).
- Admission of lack of memory (acknowledgement of gaps in memory).
- Doubts about one's own testimony.
- Self-disapproval (critical attitude to one's own behaviour).
- Forgiveness of the perpetrator of the crime (victim's statement favours the accused, or avoidance of more accusations).

e) Specific elements of the assault:
- Characteristic details of the offence (descriptions that contradict the usual beliefs about the crime).

These criteria of content can be analyzed as present or absent, or can be scored according to the strength or degree with which they appear in the statement. In any case, the presence of these aspects will favour the interpretation that the statement is true, while from their absence it cannot be inferred that it is false. With regard to the cut-off point for discriminating between statements based on truth and the rest, Steller (1989) found true statements to contain at least 7 truth criteria.

IN SEARCH OF PSYCHOLOGICAL INJURY FROM THE CRIME: THE DETECTION OF MALINGERING IN CLINICAL ASSESSMENT

Being a victim is understood as the fact of having been the object of a crime, but it also involves the whole set of consequences of the criminal act. These can be of various types: physical, economic, social or mental. Psychological injury, like memory trace, may in an expert assessment come to constitute evidence for prosecution. However, in a context such as the one we are concerned with here, the medical-legal context, it is not sufficient to diagnose a disorder or disorders: it is also necessary to rule out malingering (American Psychiatric Association, 2002). For this dual objective –clinical diagnosis and control of malingering– ordinary clinical assessment is not effective. Indeed, traditional clinical assessment has never given information on malingering (e.g., Rogers, 1997). For the measurement of psychological injury and control of malingering (hypothesis to be checked in the measurement of psychological injury caused by a crime), Arce, Fariña and Pampillón (2002) have created and validated a protocol in accordance with the responses and strategies employed by malingerers. This is based on the operative distinction between positive criteria, which validate the protocol, and negative criteria, which invalidate or mitigate its validity, indicating simulation or malingering. Positive criteria would be those not detected in the protocols of malingerers, and these are *non-avoidance of responses* and *social desirability*. Specifically, those subjects assessed by the MMPI (Minnesota Multiphase Personality Inventory) scales and who significantly refrain from responding (? Scale) and tend to give responses of social desirability do not follow the typical strategies of the malingerer, so that this should be interpreted more as an indication of truth of the protocol than as an attempt at simulation. It should be borne in mind that lack of cooperation in the assessment (non-response) had been proposed as a reliable indicator of malingering (e.g., Rogers, 1992; Lewis & Saarni, 1993; Bagby et al., 1997), but this contingency was never observed among malingerers in a forensic assessment.

The negative criteria, that is, observed in the protocols of the malingerers, were: 1) the measurement systems of MMPI, interview or others do not detect, in valid protocols, mental illness (in other words, if the measurement instruments fail to detect any mental disorder, no such disorder can be imputed in the legal context); 2) detection of malingering by the validity control scales of the MMPI and its combinations; 3) detection of some malingering strategy in the interview; and 4) lack of inter-measure agreement. The first criterion is eliminatory: if the mental disorder is not measurable, no psychological injury can be imputed as evidence. The others, in themselves, are not determinant, so that fulfilment of at least two criteria and the study of alternative hypotheses would be necessary to draw a conclusion in relation to simulation of psychological injury.

It is for these latter indicators of non-validity that we formulated the concept of *convergent invalidity*, which requires at least two totally independent indicators of for assessing a protocol as invalid. In accordance with these criteria, we drew up the following proposal for an action protocol:

a) Use of complementary and concordant systems of measurement, which presuppose the performance of different tasks involving validity control systems. Thus, it is proposed to subject the witness to a psychometric assessment, involving a symptom-recogni-

tion task, and also to apply a knowledge task, the *Clinical-Forensic Interview*. As regards the psychometric instrument to be used, the MMPI is the instrument of reference for the assessment of psychological injury in forensic practice (Butcher & Miller, 1999), but it requires a high level of comprehension on the part of the subject. Should subjects have difficulty filling out the MMPI, the SCL-90-R checklist permits the assessor to circumvent this problem, and has measures for validity control of the protocol. Therefore, we recommend this instrument as a substitute for the MMPI or, in case of doubt, as a complement. The first measure is taken through the interview to control the effect of learning the psychometric task in the knowledge task. With regard to inter-measure agreement, it must borne in mind that this will not be total. We should take into account that even test-retest measures fall short of perfection.

b) Analysis of the internal consistency of the measures: scales for control of the psychometric instruments, and, in the interview, content analysis seeking common malingering strategies. The validity control scales of the MMPI-2 (Hathaway & McKinley, 1999) with implications for the study of malingering according to this protocol are the original validity scales (no-response scales, L, F and K), the additional indicators of protocol validity (F posterior, TRIN, VRIN), and the indices that have proved effective in the detection of disorder simulation, the *F-K index* and the *inverted-V profile* (Duckworth and Anderson, 1995). If the psychometric assessment is obtained through the SCL-90-R (Derogatis, 2002), the validity scales would be PST, PSI, GSI and PSDI. As regards the interviews, these are subjected to content analysis taking as categories the strategies followed by malingerers in the interviews: avoidance of response, strange symptoms, combination of symptoms, obvious symptoms, consistency of symptoms, improbable symptoms, indiscriminate grouping of symptoms and severity of symptoms.

c) It is advisable for two assessors to separately carry out the assessment, so that inter-assessor consistency can be measured. Briefly, this safeguard serves to control for possible biases of measurement and interpretation in the assessor.

d) Study of reliability of the assessment: internal, inter-measure, inter-context (antecedents, documentary evidence, etc.) and inter-assessor consistency (Wicker, 1975).

e) Control of false positives, that is, real disorder sufferers, through the study of the subject's antecedents and general history, of the alternative hypotheses in each non-validity indicator (see Roig Fusté, 1993, Graham, 1992), and of fulfilment of the Clinical Decision Model criteria for the establishment of malingering (Cunnien, 1997).

f) Anamnesis or study of antecedents. The aim here is to reinforce the assessment with the subject's antecedents, data from his/her social context, a study of his/her behaviour, compilation of documentary evidence, other testimonies, and so on.

g) Psychological study of psychological injury. The clinical measures provide data with respect to what is legally referred to as the *biological assessment*, but also required, according to the legal demands, is the *psychological assessment*, which clarifies the relationship between the psychological injury measured and the psychological injury expected for that case.

h) Finally, the discriminant validity can also be tested. In other words, assessors can apply a measure unrelated to the case, such as on values or personality (16-PF, SIV), with the expectation of no relationship with the objective assessment, in order to rule out an attempt by the subject at manipulation of his/her image, either positively or negatively.

The resulting impression about malingering must be fitted to one of the following categories: *probable malingerer and probable non-malingerer*. It is important to avoid attempting to establish certainty (e.g., situating the impression on a scale of several points), as this creates confusion in decision-makers (e.g., sentence of the Spanish Supreme Court, 29 October, 1981, RA 3902), and it is crucial to use probabilistic terms, since psychological assessment is subject to error.

THE GLOBAL EVALUATION SYSTEM

The *Global Evaluation System* (GES) is structured around 9 tasks which we shall briefly describe and explain below: obtaining the statement, repeating the statement, checking of the statements obtained in the course of the judicial process, content analysis of the statements, reliability analysis of the measures, measurement of clinical effects of the traumatic event, assessment of statements from persons involved, analysis of personality and capacities of those involved, and finally, implications for presentation of the report. The tasks to be performed are mediated by the case to be assessed (e.g., if in a given

case it is not possible to assess the accused, this phase is not implemented). The phases in the most complete version of the system are as follows:

a) *Obtaining the statement (memory trace).* For the forensic psychological procedure for assessment of statements and psychological injury to be productive, reliable and valid, assessors need instruments for obtaining the statement and measuring the clinical condition that permit their subsequent analysis. Therefore, the statements must be obtained, depending on whether the subjects are adults, minors or disabled persons, through the following procedures: Improved Cognitive Interview (Fisher & Geiselman, 1992), Memorandum of Good Practice (Bull, 1997), or the Forensic Interview for the Disabled (Arce, Novo & Alfaro, 2000). Psychological injury in the *knowledge task* is measured through the *Clinical-Forensic Interview* (Arce & Fariña, 2001; Arce, Fariña & Freire, 2002; Arce, Pampillón & Fariña, 2002).

b) *Repetition of obtaining the statement.* The methods proposed are based on a single measure of the memory trace. However, with a single measure, we lose the possibility of analyzing the temporal consistency of the statement (below we discuss the validity of the other statements obtained in the course of the judicial process). In other words, we leave out one of the forms of checking the validity of the information: temporal or intra-witness consistency (e.g., Wicker, 1975; Schum, 1977). Likewise, legal doctrine has defined the reliability of a testimony according to opportunity criteria (opportunity to observe, etc.), bias (control of possible interests), temporal consistency, plausibility, inter-witness consistency and credit (Schum, 1977). Also, our own jurisprudence (e.g., sentence of the Spanish Supreme Court, 29 April, 1997) establishes, when the testimony of the victim is the sole or central evidence for the prosecution, that the testimony must display the following three characteristics: absence of subjective incredibility, some peripheral corroboration of an objective nature, and persistence over time without ambiguities or contradictions. In this line, sentences have already been pronounced that annul the evidence value of content analysis (CBCA and SVA) of statements based on a single statement (e.g., AP, Pontevedra, Sección 6ª, 21 January, 2004). In sum, both scientific methodology and legal doctrine and jurisprudence demand more than one statement for the study of temporal consistency. In this regard, it has been found that the repetition of obtaining the statement need not contaminate the data from an interview not contaminated from outside (e.g., Campos & Alonso-Quecuty, 1999), as is the case for the protocols for obtaining the statement mentioned previously. Consequently, in the first measure the assessor must by no means interrogate the subject, employing solely the reconstruction of contexts, free recall, change of perspective and reverse-order recall. Interrogation, where necessary, is to be left for the second measure, so as not to contaminate the memory of events with the interrogation. From a second measure, the assessors obtain a consistency analysis, which, according to the Undeutsch hypothesis (1967, p. 125), should be understood as a function of the centrality/periphery of the contradictory material. It should be pointed out here that the contradiction is only relevant if it affects central details for the act of judgement. Inconsistency in peripheral information or the omission of certain information is only important if that information is crucial to the construction of a real event. In order to leave room for interferences (theory of the interference of forgetting), the entry of new information (constructive hypothesis of forgetting) and the forgetting curve, we estimate the time that should elapse between interview and interview at over 1 week (but not much more). We establish three axioms with respect to this. First, since the criminal act constitutes a *stressful life event*, the *obsolescence effect* will be weaker (in reference to the testimony of both plaintiff and defendant, and contiguous with the facts). Second, a *theory of rationality* on the part of the fabricator, so that the lie is planned, learned and, by extension, consistent in time; hence, the lie will not be mediated by post-event interference and information (*constructive hypothesis*). Here, it is essential to obtain the first statement in the free narrative format, without any kind of interrogation, to avoid letting in post-event information that the subject would fit into the new reconstruction. Interrogation would only take place after obtaining the second statement in free narrative. Third, the subject who is telling the truth narrates images, so that the description of the facts, though quite similar, will be constructed differently, as it does not correspond to episodic schemata. In sum, and in free recall format, the true statement will be less consistent, and although the event is the same, the narration will be significantly different, in

terms of both its retrieval and its content (omissions, elicitation of events other those under investigation but related to them, inconsistency in peripheral information, retrieval of new information of little relevance to the events). For their part, simulating subjects narrate learned stories, so that they repeat them more or less the same each time, guided by an episodic schema. It is important to bear in mind that this second statement should always be considered from the perspective that does not contribute significantly to secondary victimhood.

c) *Checking of the statements obtained in the course of the judicial process*. Similarly, an analysis is made, according to the procedure of the validity study in SRA and SVA, of the other statements made in the course of the judicial process (e.g., indictment, inquiry). However, the value of these is relative. It should be borne in mind that many of them are transcriptions of what the witness has said, so that they do not reliably reflect the testimony. Moreover, the type of interrogation may have influenced the response. In this regard, one should bear in mind, in line with SVA, the effects on statement validity of interview characteristics (types of question formulated and suitability of the interview) and motivation (motives, context and pressure). For example, in the case of interrogations of children we have found many expressions and concepts of which the child, when asked, does not know the meaning (e.g., in the minor's statement there appears the expression *semen came out*; if the child, on being asked what *semen* is, does not know, then this expression does not belong to his/her statement). In turn, statements often refer to expressions (e.g., he raped me), rather than to narrations of events, so that the reliability and validity cannot be checked. Thus, the lack of consistency of the statements given to the expert investigators and others included in the judicial proceedings has a quite relative value. Where appropriate, it should be explained that this lack of consistency is not relevant for analysis of the plausibility of the statement. Furthermore, it is important to exercise more caution than we might initially think on considering confessions by the accused, and even more so, incriminations in exchange for benefits accruing to the informer. The source of bias can be found in the interrogations. Thus, the usual techniques for obtaining a confession are based on strategies such as: threats; attribution of responsibility to external causes, such as provocation by the victim; minimization of the seriousness of the crime; or the development of a personal relationship with the suspect (i.e., the typical "good cop, bad cop" strategy, with two interviewers, one hostile and the other friendly and protective). Finally, the strategy based on the Prisoner's Dilemma for obtaining the statement may lead to either cooperation strategies or competition strategies that distort the expression of the testimony (e.g., Kelley & Stahelski, 1970). In this regard, a decision by the US Supreme Court (Miranda v. Arizona, 1966) declared this type of interrogation as coercive.

d) *Content analysis of statements referring to the events.* Content analysis of the statements addresses two dimensions: the validity and the reliability of the testimony. According to the *Global Evaluation System*, the first task of the expert assessment consists in estimating the validity of the statement, not as judicial evidence per se –which is the business of the Judicial System–, but as evidence whose reliability is to be analyzed. In this regard, there are two types of potential attack on validity. First, the statement may be of insufficient length to be subjected to a reality analysis (Raskin & Steller, 1989); and second, the statement may be considered invalid as evidence on the basis of the validity criteria of the SRA and SVA (e.g., lack of internal consistency; lack of external consistency with other robust or incontrovertible evidence, such as that obtained by experts during the course of the judicial process; statement inconsistent with a previous one; lack of persistence in statements; inconsistency with the laws of science and nature) and, in the case of minors, indicators that limit the validity (indicators of suggestibility, inappropriateness of affect, inappropriateness of language and knowledge). If the evidence is deemed invalid, it is concluded that the statements do not constitute admissible or sufficient evidence; if it is deemed valid, the reliability (consistency with criteria of reality) of the statements is analyzed. As a categorial list of reference we use the categories of the CBCA. This analysis procedure, created in principle for the testimony of minors who were victims of sexual assault, is equally effective with adults (Landry & Brigham, 1992; Zaparnuik, Yuille & Taylor, 1995; Spörer, 1997; Vrij, Edward, Roberts & Bull, 1999), in sequences of measures, and in cases other than those of sexual assault (Porter & Yuille, 1996; Spörer,

1997; Arce, Fariña & Freire, 2002). In these new contexts, obviously not all the categories are productive. Thus, Landry and Brigham (1992) restrict use to 14 categories with adults, since three are applicable only to children (incomprehension of details related to accuracy; forgiveness of the perpetrator of the crime; and details characteristic of the offence), while another two (lack of structure and related external associations) were not productive. However, we (Arce, Fariña & Freire, 2002) found that the category *forgiveness of the perpetrator of the crime* was productive, in statements by adults, both for sexual assault and threats –that is, the productivity is subject to a context effect. In sum, all the criteria should in principle be considered in the analysis, since productivity depends on type of case, peculiarities of the action under examination and interviewee's sociodemographic profile. In turn, the combination of the CBCA and RM criteria is possible and effective, as their effects can be summed (Spörer, 1997; Vrij et al., 1999). Specifically, the combination of the two assessment systems, through the addition to the CBCA of the RM criteria *perceptual information* and *cognitive operations* (Vrij, 2000), slightly improves the reliability of the system. Thus, these two new criteria can be added to those of the CBCA. This procedure can be applied in repeated measures (see the hypotheses to be tested in the section *repetition of obtaining the statement*).

e) *Reliability analysis of the measures*. The original systems of statement content analysis constitute semi-objective techniques because they only examine the reliability and validity of the instruments, and do not contain procedures for control of the specific measure, i.e., the expert assessor's measure. With a view to addressing this methodological shortcoming and approaching an objective system, we propose a method that makes it possible to verify the reliability of the measure through the analysis of inter- and intra-measure, inter-assessor and inter-context consistency (Wicker, 1975). Inter-context reliability is addressed through recourse to a trained assessor who has been effective and consistent in other, previous contexts, that is, in previous expert investigations. By using two assessors (at least one of whom has been trained and showed reliability in previous assessments) performing the tasks separately, it is possible to make an assessment of inter-assessor consistency. As a statistical tool for the analysis of inter-assessor consistency, we propose the Agreement Index [AI= Agreements/(agreements+disagreements)], which is more restrictive than the kappa values, taking as cut-off point a value of .80 (Tversky, 1977). In other words, the results are only considered reliable if two assessors, separately, agree on more than .80 of the total assessments in each analysis category. Inter- and intra-measure consistency are checked by means of: internal consistency of the measures (e.g., the validity scales of the MMPI, the statements or the study of malingering strategies in the clinical interview); consistency between different measures (e.g., agreement between MMPI and clinical interview, between statements over time); and consistency –i.e., complementariness or its lack (one may present indicators of truth and the other indicators of falsity, or none at all)– between the assessments obtained for the plaintiff and the defendant.

f) *Measurement of clinical effects of the traumatic event.* The criminal act causes a series of injuries to the victim that are basically of a physical, psychological and economic nature. The psychological damage constitutes the so-called psychological injury of the crime and, as such, can be adduced as evidence for the prosecution. In relation to the assessment of psychological injury and the subsequent judicial evidence, criminal acts (e.g., lesions, breaking and entering, abuse, sexual abuse, kidnapping) can produce a psychological response corresponding to a diagnosis of post-traumatic stress disorder (PTSD) (Blanchard & Hickling; 2004, Echeburúa & Corral, 1998; Echeburúa, Corral, Zubizarreta & Sarasúa, 1995). Therefore, the measurement of PTSD is crucial for the detection of psychological injury. Special care should be exercised with indirect measures of PTSD (e.g., hypochondria, hysteria, depression, anxiety, dysthymia, social isolation, social maladjustment), which can serve as enhancers of PTSD, but are not substitutes for it. Furthermore, it is necessary to rule out causes other than the criminal act. For example, the combination of a process of divorce or separation with abuse may make it difficult to distinguish the source of the disorder, since the two contingencies produce similar psychological injury. In any case, the expert assessor must take into account the following maxims: not all criminal acts produce PTSD in the victim; and the absence of PTSD does not imply that the assault did not take place. At the

same time, once the psychological injury has been identified, it is necessary to check whether it is real or simulated, and to this end assessors can use the protocol previously described for the measurement of psychological injury with control of malingering.

g) *Assessment of statements from persons involved.* Although in principle the techniques of content analysis of statements and assessment of psychological injury were designed for the assessment of the victim or plaintiff's testimony, the same procedure can be applied to the defendant, thus making possible a study of the two versions together. The inquisitorial justice procedure, as employed in Spain, permits this confrontation (though this would not be the case with an adversarial system). With this procedure we can obtain an estimation of convergent validity.

h) *Analysis of personality and capacities of those involved.* The study of the personality of the actors can be crucial to an explanation of the accusation itself, the assault or any mental disorder of the defendant with relevant judicial implications; in other words, where applicable, the imputability of the accused is examined [see Arce, Fariña & Pampillón (2002) for a description of how the study of imputability is carried out]. Since a clinical assessment is not sufficient in the forensic context, the clinical assessment protocol with control of malingering (Arce, Fariña & Pampillón, 2002) is followed. Actors' cognitive capacities are measured by means of the corresponding Wechsler Scale, and as a source of contrast or for samples with language difficulties, poor education or from other countries, the non-verbal intelligence test TONI-2 (Brown, Sherbenou & Johnsen, 1995) is taken. Reliability of this last-named measure is checked through correspondence of responses with the difficulty gradient of the questions and inter-measure consistency. We use the assessment of cognitive capacities to rate the capacity to testify, and, where appropriate, to indicate their effects on criminal responsibility.

i) *Implications for presentation of the report.* The system of statement credibility in 5 response categories, as proposed in SVA, does not meet the requirements of the Spanish Judicial System. The Supreme Court demands complete certainty, not merely high probability (e.g., sentence of the Spanish Supreme Court, 29 October, 1981, RA 3902). However, all measures, and particularly psychological ones, are subject to error, so that we should acknowledge this, but refraining from establishing degrees of certainty which, in accordance with the considerations of the Supreme Court, lead only to greater confusion. Thus, the most appropriate categories would be *probably true*, *probably untrue* and, where applicable, *indeterminate* (interested forensic psychologists can obtain from the authors an expert assessment format based on the GES).It should also be borne in mind that the system is more robust in the identification of truth than of lies. Likewise, it is advisable not to make a description of events based on phrases, but rather on complete actions, since the procedure validates events, and not isolated parts. Therefore, in no case is it recommended to identify the alleged perpetrator, as the procedure does not validate this point.

FINAL CONSIDERATIONS

The reliability of the entire procedure is ultimately the responsibility of the interviewer/assessor. It is for this reason that the intervention be carried out by well trained and experienced professionals with a high capacity for objectivity (Alonso-Quecuty, 1993). Thus, exhaustive training is essential. This should include: a) training in all forms of obtaining all types of information (a procedure can be seen in Fisher et al., 1987); b) training in statement analysis [a structured programme can be found in Köhnken (1999)]; c) training in the assessment of personality and psychological injury, not for clinical, but for forensic purposes (see Arce, Fariña & Freire, 2002; Arce, Fariña & Pampillón, 2002; Echeburúa, Corral & Amor, 2002; Rogers, 1997); d) training in the detection of malingering (see the steps to be followed in Arce, Fariña & Pampillón, 2002); and e) first forensic assessments to be performed in the company of an expert investigator with experience. Finally, our experience suggests that the material used in training in content analysis and clinical assessment should be real, rather than simulated, since the task executed in the two contexts is different, and the effectiveness of the procedure also (Vrij, 2000). The Forensic Psychology Unit at the University of Santiago de Compostela periodically organizes training courses in these techniques.

ACKNOWLEDGEMENTS

This research was financed by the *Secretaría Xeral de Investigación e Desenvolvemento* (General Secretariat for Research and Development) of the Galician Regional

Government, within the *Excelencia Investigadora* (Research Excellence) project, Code: PGIDIT03CS037401PR.

REFERENCES

Alonso-Quecuty, M. L. (1993). Interrogando a testigos, víctimas y sospechosos: La obtención de información exacta. In M. Diges, & M. L. Alonso-Quecuty (eds.), *Psicología forense experimental* (pp. 85-98). Valencia: Promolibro.

American Psychiatric Association, (2002). *Manual diagnóstico y estadístico de los trastornos mentales* (IV Edición-Texto Revisado). Barcelona: Masson.

Arce, R. & Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de delitos: La entrevista clínico-forense*. Unpublished manuscript, Universidad de Santiago de Compostela.

Arce, R. & Fariña, F. (2002). *El Sistema de Evaluación Global para la evaluación de la credibilidad y el engaño en el testimonio judicial*. Unpublished manuscript, Universidad de Santiago de Compostela.

Arce, R. & Fariña, F. (2005). El Sistema de Evaluación Global (SEG) de la credibilidad del testimonio: Hacia una propuesta integradora. In R. Arce, F. Fariña & M. Novo (eds.). *Psicología jurídica* (pp. 101-118). Santiago de Compostela: Consellería de Xustiza, Interior e Administración Local.

Arce, R. & Fariña, F. (in press). Evaluación de la credibilidad y de la huella psíquica en el contexto penal. In Consejo General de Poder Judicial (ed.), *Psicología del testimonio y valoración de la prueba pericial*. San Sebastián: Centro de Documentación Judicial.

Arce, R., Carballal, A., Fariña, F. & Seijo D. (2004). Can mock battered women malinger psychological evidence in a recognition task? In A. Czerederecka, T. Jaskiewic- Obydzinska, R. Roesch & J. Wójcikiewicz (eds.), *Forensic psychology and law. Facing the challenges of a changing world* (pp. 327-336) Kraków: Institute of Forensic Research Publishers.

Arce, R., Fariña, F. & Freire, M. J. (2002). Contrastando la generalización de los métodos empíricos de detección del engaño. *Psicologia: Teoria, Investigação e Prática*, 7(2), 71-86.

Arce, R., Fariña, F. & Real, S. (2000). The assessment of information integration theory and confirmatory bias hypothesis on judicial proceedings. A case of rape and murder. In A. Czerederecka, T. Jaskiewic-Obydzinska & J. Wójcikiewicz (eds.), *Forensic psychology and law* (pp. 296-303) Kraków: Institute of Forensic Research Publishers.

Arce, R., Fariña, F. & Seijo, D. (2003). Laypeople's criteria for the discrimination of reliable from non-reliable eyewitnesses. In M. Vanderhallen, G. Vervaeke, P. J. Van Koppen & J. Goethals (eds.), *Much ado about crime* (pp. 105-116). Brussels: Uitgeverij Politeia NV.

Arce, R., Fariña, F., Carballal, A. & Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación. *Psicothema*, 18.

Arce, R., Fariña, F., Novo, M. & Seijo, D. (2001). Judges' decision-making from within. In R. Roesch, R. R. Corrado & R. J. Dempster (eds.), *Psychology in the courts: international advances in knowledge* (pp. 195-206). New York: Routledge.

Arce, R., Novo, M. & Alfaro, E. (2000). La obtención de la declaración en menores y discapacitados. In A. Ovejero, M. V. Moral & P. Vivas (eds.), *Aplicaciones en psicología social* (pp. 147-151). Madrid: Biblioteca Nueva.

Arce, R., Pampillón, M. del C. & Fariña, F. (2002). Desarrollo y evaluación de un procedimiento empírico para la detección de la simulación de enajenación mental en el contexto legal. *Anuario de Psicología*, 33(3), 385-408.

Bagby, R. M., Buis, T. & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: replication and extension. *Psychological Assessment*, 7, 84-92.

Bagby, R. M., Rogers, R. Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., Schuller, D. R. & Seeman, M. V. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, 68(3), 650-664.

Bauer, P. J. & Mandler, J. M. (1990). Remembering what happened next: Very young children's recall of event sequences. In R. Fivush, & J. A. Hudson (eds.), *Knowing and remembering in young children* (pp. 9-29). New York: Cambridge University Press.

Blanchard, E. B. & Hickling, H. J. (2004). What are the psychosocial effects of MVAs on survivors? In E. B. Blanchard, E. J. Hickling (eds.), *After the crash: Psychological assessment and treatment of survivors of motor vehicle accidents* (2ª Ed., pp. 57-97). Washington, DC: American Psychological Association.

Bower, G. (1967). A multicomponent theory of memory trace. In K. W. Spence & J. T. Spence (eds.), *The psychology of learning and motivation* (Vol. 1) (pp. 229-325). New York: Academic Press.

Bower, G. H. & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247, 44-48. Brown, L., Sherbenou, R. & Johnsen, S. (1995). *Toni-2*. Madrid: TEA Ediciones.

Bull, R. (1995). Interviewing people with communicative disabilities. In R. Bull & D. Carson (eds.), *Handbook of psychology in legal contexts* (pp. 247-260). Chichester: John Wiley and Sons.

Bull, R. (1997). Entrevistas a niños testigos. In F. Fariña & R. Arce (eds.), *Psicología e investigación judicial* (pp. 19-38). Madrid: Fundación Universidad Empresa.

Butcher J. N. & Miller, K. B. (1999). Personality assessment in personal injury litigation. In A. K. Hess & I. B. Weiner (eds.), *The handbook of forensic psychology* (2ª Ed., pp. 104-126). New York: John Wiley and Sons.

Campos, L. & Alonso-Quecuty, M. L. (1999). The cognitive interview: Much more than simply "try again". *Psychology, Crime and Law*, 5(1-2), 47-59.

Cunnien, A. J. (1997). Psychiatric and medical syndromes associated with deception. In R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 23-46). New York: Guilford Press.

Dammeyer, M. D. (1998). The assessment of child sexual abuse allegations: Using research to guide clinical decision making. *Behavioral Sciences and the Law*, 16, 21-34.

Davies, G., Tarrant, A. & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology*, 80, 415-429.

Derogatis, L. R. (2002). *SCL-90-R. Manual*. Madrid: TEA Ediciones.

Duckworth, J. & Anderson, W. P. (1995). *MMPI and MMPI-2: Interpretation manual for counsellors and clinicians.* Bristol: Accelerated Development.

Echeburúa, E. & Corral, P. (1998). *Manual de violencia familiar*. Madrid: Siglo XXI.

Echeburúa, E., Corral, P. de & Amor, P. J. (2002). Evaluación del daño psicológico en víctimas de crímenes violentos. *Psicothema*, 14, 139-146.

Echeburúa, E., Corral, P. de, Zubizarreta, I. & Sarasúa, B. (1995). *Trastorno de estrés postraumático crónico en víctimas de agresiones sexuales*. A Coruña: Fundación Paideia.

Fisher, R. P. & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interview*. Springfield: Charles C. Thomas.

Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M. & Warhaftig, M. L. (1987). Enhancing eyewitness memory: Refining the cognitive interview. *Journal of Police Science and Administration*, 15, 291-297.

Freed, D. M., Corkin, S., Growdon, J. H. & Nissen, M. J. (1989). Selective attention in Alzheimer's disease: Characterizing cognitive subgroups of patients. *Neuropsychologia*, 27, 325-339.

Gil, R. (1999). *Neuropsicología*. Barcelona: Masson.

Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M. & Kuhn, J. (1997). Children's reactions to and memory for a stressful event: Influences of age, anatomical dolls, knowledge, and parental attachment. *Applied Developmental Science*, 1, 54-75.

Graham, J. R. (1992). *Interpretation of MMPI-2. Validity and clinical scales*. Bruges, workshop.

Hart, R. P. & O'Shanick, G. J.(1993). Forgetting rates for verbal, pictorial, and figural stimuli. *Journal of Clinical and Experimental Neuropsychology*, 15, 245-265.

Hathaway, S. R. & McKinley, J. C. (1999). *MMPI-2. Inventario multifásico de personalidad de Minnesota-2. Manual*. Madrid: TEA Ediciones.

Home Office and The Department of Health (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: HMSO.

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.

Kelley, H. H. & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 16, 66-91.

Köhnken, G. (1999). *Assessing credibility*. Pre-conference of the EAPL Programme of Applied Courses, Dublin.

Köhnken,G., Milne, R., Memon, A. & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, 5, 3-27.

Landry, K. L. & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, 16, 663-676.

Lewis, M. & Saarni, C. (1993). *Lying and deception in everyday life*. New York: Guilford Press.

Loftus, E. F., Korf, N. L. & Schooler, J. W. (1988). Misguided memories: Sincere distortions of reality. In J. C. Yuille (ed.), *Credibility assessment* (pp. 155-172). Dordrecht: Kluwer.

Mantwill, M., Köhnken, G. & Ascherman, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology, 80*, 68-78.

Memon, A. & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation and practical implications. *Journal of Community and Applied Social Psychology*, 1, 291-307.

Memon, A., Cronin, O., Eaves, R. & Bull, R. (1993). The cognitive interview and the child witness. In N. K. Clark & G. M. Stephenson (eds.), *Issues in criminology and legal psychology: Vol. 20. Children, evidence and procedure*. Leicester: British Psychological Society.

Memon, A. Cronin, O., Eaves, R. & Bull, R. (1996). An empirical test of mnemonic components of cognitive interview. In G. M. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (eds.), *Psychology and law: Advances in research* (pp. 135-145). Berlin: Walter de Gruyter.

Memon, A., Wark, L., Bull, R. & Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology, 88*, 179-198.

Memon. A. & Bull, R. (1999). Handbook of the psychology of interviewing. Chichester: John Wiley and Sons.

Mira, J. J. (1989). *Estudios de psicología en ambientes procesales: Un análisis de metamemoria*. Doctoral Thesis, Universidad Autónoma de Madrid.

Miranda v. Arizona, 384 U.S. 436 (1966).

Ostrom, T. M., Werner, C. & Saks, M. J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 36, 436-450.

Peña-Casanova, J. (1990). *Programa integrado de exploración neuropsicológica. Test Barcelona*. Barcelona: Masson.

Piñeiro, A. (2005). Criterios empíricos de credibilidad y profesionales de la justicia. In R. Arce, F. Fariña & M. Novo (eds.). *Psicología jurídica* (pp. 213-221). Santiago de Compostela: Consellería de Xustiza.

Porter, S. & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues in the interrogation context. *Law and Human Behavior*, 20(4), 443- 458.

Raskin, D. C. & Steller, M. (1989). Assessing credibility of allegations of child sexual abuse: Polygraph examinations and statement analysis. In H. Wegener, F. Lösel & J. Haisch (eds.), *Criminal behaviour and the justice system: Psychological perspectives* (pp. 290-302). Heidelberg: Springer-Verlag.

Rogers, R. (1992). *Structured interview of reported symptoms*. Odessa, Fl.: Psychological Assessment Resources.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 398-426). New York: Guilford Press.

Rogers, R. & Mitchell, C. N. (1991). *Mental health experts and the criminal courts: A handbook for layers and clinicians*. Scarborough, ON: Thomson Professional Publishing Canada.

Roig-Fusté, J. M. (1993). *MMPI y MMPI-2 en la exploración de la personalidad*. Barcelona: Agil Offset.

Sallmann, P. & Willis, J. (1984). *Criminal justice in Australia*. Melbourne: Oxford University Press.

Schum, D. A. (1977). The behavioral richness of cascaded inference models: Examples of jurisprudence. In N. J. Castellan, D. B. Pisoni & G. R. Potts (eds.), *Cognitive theory* (vol. 2). Hillsdale, N.J.: LEA.

Spörer, S. L. (1997). The less traveled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology,* 11, 373-397.

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (ed.), *Credibility assessment*. (pp. 135-154). Dordrecht: Kluwer Academic Publishers.

Steller, M. & Köhnken, G. (1994). Análisis de declaraciones basados en criterios. In D. C. Raskin (ed.), *Métodos psicológicos en la investigación y pruebas criminales* (pp. 217-245). Bilbao: Desclée de Brouwer (Orig. 1989).

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 353-370.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen. In U. Undeutsch (ed.), *Handbuch der psychologie, Vol. II: Forensische psychologie* (pp. 26-181). Göttingen: Verlag für Psychologie.

Undeutsch, U. (1988). The development of statement reality analysis. In J. Yuille (ed.), *Credibility assessment* (pp. 101-119). Dordrecht: Kluwer.

Vrij, A. (2000). *Detecting lies and deceit*. Chichester: Wiley.

Vrij, A. & Winkel, F. W. (1996). Detection of false statements in first and third graders: The development of a nonverbal detection instrument. In G. Davies, S. Lloyd-Bostock, M. McMurray & C. Wilson (eds.), *Psychology, Law, and Criminal Justice* (pp. 221-230). Berlin: Walter de Gruyter.

Vrij, A., Edward, K., Roberts, K. P. & Bull, R. (1999, July). *Detecting deceit via criteria based content analysis, reality monitoring and analysis of non verbal behavior*. The First Joint Meeting of the American Psychology- Law Society and the European Association of Psychology and Law, Dublin.

Wechsler, D. (1976). *WAIS: Escala de inteligencia para adultos (2ª ed.).* Manual. Madrid: TEA Ediciones.

Weick, K. E. (1985). Systematic observational methods. In G. Lindzey & E. Aronson (eds.), *The handbook of social psychology* (vol. 1, pp. 567-634). Hillsdale, N.J.: LEA.

Wicker, A. W. (1975). An application of a multipletrait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.

Winograd, E., Smith, A. D. & Simon, E. W. (1982). Aging and the picture superiority effect in recall. *Journal of Gerontology*, 37, 70-75.

Zaparniuk, J. Yuille, J. C. & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, 18, 343-352.

Papeles del Psicólogo, 2005. Vol. 26, pp. 78-91

IS THE LIAR CAUGHT SOONER THAN THE CRIPPLE? POPULAR WISDOM VERSUS SCIENTIFIC KNOWLEDGE ON THE NON-VERBAL DETECTION OF DECEPTION

Jaume Masip

University of Salamanca

There are a series of popular beliefs about the behavioural indicators of deception that are not supported by empirical research. A number of "self-help" books are contributing to the spreading of these beliefs. In this article, several decades of psychological and communication research on the non-verbal detection of deception are reviewed. Contrary to the claims of "self-help" books and to the tenets of popular wisdom, detecting deception from behavioural cues is extremely difficult, there are almost no behavioural cues to differentiate between truths and lies, their meaning and usefulness depend on a number of contextual variables, and training programmes have yielded only very limited improvements in accuracy. In view of the misleading content of certain popular books and the serious consequences of wrong credibility judgments in a number of contexts, it is necessary to dispel the existing myths about the non-verbal detection of deception, providing instead valid and scientifically tested information.

Existe una serie de creencias populares sobre los indicadores conductuales del engaño que no se ven corroboradas por la evidencia empírica. Determinados libros "de autoayuda" contribuyen a la difusión de las mismas. En este trabajo se revisan varias décadas de investigación en psicología y comunicación sobre la detección no-verbal del engaño. Al contrario de lo que propugnan los libros "de autoayuda" y de lo que sostiene la sabiduría popular, detectar la mentira a partir del comportamiento no-verbal es extremadamente difícil, apenas sí existen claves conductuales que permitan discriminar entre verdades y mentiras, su significado y poder de discriminación varían en función de diversas variables contextuales, y la eficacia de los programas de entrenamiento es muy limitada. Frente a las cuestionables afirmaciones de determinados libros populares y dadas las graves consecuencias que en ciertos ámbitos pueden tener los juicios de credibilidad erróneos, es necesario desmontar los falsos mitos existentes sobre la detección no-verbal de la mentira, sustituyéndolos por información más válida y científicamente contrastada.

Social psychology has examined on numerous occasions the relationship between psychological knowledge and common sense (e.g., Garrido, Herrero & Masip, 2004; Teigen, 1986; see the discussion by Kelley, 1992). As many authors have pointed out (e.g., Myers, 1999/2000), social psychology is criticized for studying things that everyone already knows –that are "common sense" (Kelley, 1992; Schlesinger, 1949). Such criticism, however, is almost always made after the critic has been given the correct answer ("I already knew that!"); in general people do not find it so easy to come up with such an "obvious" response themselves (Kelley, 1992; Lazarsfield, 1949).

Correspondence: Jaume Masip. Departamento de Psicología Social y Antropología. Universidad de Salamanca. Facultad de Psicología. Avda. de la Merced, 109-131. 37005 Salamanca. Spain. E-mail: jmasip@usal.es

There are two areas in which I have maintained a professional interest for some years where the distance between common sense and scientific evidence is particularly notable. These are the areas of non-verbal communication and the detection of deception. Due in all probability to their intrinsic attraction, both topics stimulate the popular imagination, giving rise to the most outlandish "theories" and views, which all too often soon become widely accepted among the general public. The dissemination of such ideas is usually helped by the contribution of an endless chain of opportunist books, misleadingly called "self-help" manuals, or similar, and often written by people with suspicious qualifications who take advantage of the ingenuity of the reader, probably with financial motives, and therefore in an entirely unethical manner. I should stress that I am not dismissing *all* self-help books. Undoubtedly, some are the work of reputable researchers and written with the utmost scientific

rigour. But a large part of these types of book are based on the naive and untested beliefs of their authors, rather than on scientific knowledge about the subject. This does little to help the dissemination of authentic science, nor does it contribute to the self-help that interested readers are seeking (which is why the term "self-help" is in inverted commas here). The curious thing is that while readers' lack of familiarity with the field is what leads them to turn to such books, it is precisely such ignorance that prevents them appreciating the scarcity of their scientific value, thus making them vulnerable to the misinformation they impart.

In more specific terms, with regard to the field of non-verbal communication (or "body language", as it is often called in these books) it turns out that, as in many other areas of psychology, everyone "knows" about the topic, and has an opinion on it, daring to deny, relativize or qualify what the true experts say. It is as though popular stereotypes had more value than scientific knowledge obtained through rigorous and well-established procedures. It is not uncommon to come across advertisements making remarkable claims about courses on "successful communication", or with sensationalist titles such as "body language for salesmen" or the like, offered by consultants or other organizations from outside our field of specialization, and delivered by people whose lack of qualifications in psychology or interpersonal communication makes them wholly unsuitable for such a task. It would be absurd for a psychologist to consider giving a course on law, economics or engineering. Nevertheless, certain lawyers, economists, engineers and others from outside the psychology profession do not hesitate to consider themselves qualified to plunge unhesitatingly into the field of psychology, communication and related social sciences to impart specialized "knowledge". In my opinion, this is nothing short of professional encroachment.

Such a state of affairs is unlikely to lead to anything but the spread of false beliefs about the meaning of behaviour, and to a distorted image of non-verbal behaviour as "child's play" to interpret, with gestures of unmistakeable meaning and totally independent of context. Thus, for example, many people believe that crossing the legs or arms clearly signifies that the person is not psychologically "open" to the other, that jutting out the chin is a sign of dominance, and so on. It is difficult not to smile at the ingenuousness of such beliefs, which reflect a series of often erroneous implicit theories rather than authentic scientific knowledge on the topic.

A good example of the dissemination of this type of belief is the well known *Body Language*, by Allan Pease (1981/1988). The author, a commission salesman, first became interested in "body language" after attending a seminar by the anthropologist Ray Birdwhistell in 1971. It is unfortunate that, in his book, Pease does not do justice to Birdwhistell's unquestionable scientific reputation, despite the deceptive claim in the foreword that "in writing this book, I have summarised many of the studies by the leading behavioural scientists" (Pease, 1988, p. 9).

As if the spread of false beliefs "disguised" as scientific knowledge by the unqualified were not already harmful enough, the matter borders on the scandalous when those spreading such ideas are supposedly professionals. Paolo Abozzi, the self-styled director of the so-called *Centro di Comunicazione Integrale* in Rome, and who claims to have training in communication and hypnosis (see http://digilander.libero.it/magopaolo/PAOLO%20ABOZZI.html), is the author of, among other works, *The Interpretation of Gestures* (Abozzi, 1996/1997). This book is of a similar nature to that of Pease, while the *Centro di Comunicazione Integrale*, despite its grand-sounding name, is actually not a research centre at all, but rather an institution that gives courses and produces videos on hypnosis, graphology, neurolinguistic programming and similar subjects (http://digilander.libero.it/magopaolo/index2.html). The danger involved in the spread of false knowledge by supposed professionals resides in the well known influence of the credibility of the source on persuasion (Kruglansi et al., 2005). The ingenuous client is likely to consider such information as true since it is provided by an "expert" in the field, believing blindly all the claims and following all the recommendations made. This can lead to erroneous decisions with serious consequences in interpersonal, work or judicial contexts.

The second area to which I referred above is that of the detection of lying or deception. Being just as "intriguing" as the field of non-verbal behaviour, it is threatened by the same dangers. These dangers are represented in this context by, for example, diverse techniques or procedures developed by seasoned police or military officers whose professional experience in situations where lying is frequent lends them a certain degree of popular credibility[1]. But the fact of a professional's *experience* does not necessarily imply that he or she is an *expert* (see, with specific reference to the field of non-verbal detection

of deception, the works of DePaulo & Pfeiffer, 1986; Garrido, Masip & Herrero, 2004; Meissner & Kassin, 2002; or Strömwall, Granhag & Hartwig, 2004). Consequently, their recommendations may be mistaken. The boom in the use of instruments such as voice stress analyzers (Masip, Garrido & Herrero, 2004) or procedures such as the SCAN Technique (Masip, Garrido & Herrero, 2002a) constitute clear examples of this. Developed by experienced professionals from the field of security, such devices and procedures enjoy considerable popularity in applied contexts, due partly to their creators' profession and partly to the powerful marketing mechanisms at their service. Nevertheless, their true utility for detecting lies has been seriously called into question by empirical research. The risk is, once more, the potentially serious consequences of the use of the erroneous information provided. If the myth that voice stress analyzers or the SCAN Technique are valid and reliable instruments or procedures is well-rooted in society, courts and judges are more likely to admit the evidence obtained with them in trials. But if in reality these techniques cannot discriminate adequately between truth-tellers and liars, we may be unfairly convicting innocent suspects, while the truly guilty go free (see, on this topic, the report by the National Research Council, 2003, in reference to the use of the polygraph).

But if the problem is already a considerable one taking non-verbal behaviour and lie detection separately, it is not surprising that the situation is bleak when it comes to the detection of lies on the basis of non-verbal behaviour. A few years ago I saw an advertisement in a catalogue for a book by one David Lieberman (1998), entitled *Never be lied to again*. I ordered it, albeit with open scepticism given the sensationalist nature of the title and the fact that the author was totally unknown to me (he was clearly not among the relevant researchers in this field). The book, subtitled "how to find out the truth in 5 minutes or less in any conversation or situation", contains absolutely no information of any scientific or practical value, but rather a collection of absurd pieces of advice that are totally misleading for the reader. The most outrageous aspect of the case is the fact that the letters "Ph.D." appear on the cover and spine of the book alongside the author's name. Likewise, the notes on the dust jacket extol the supposed professional virtues of Dr. Lieberman. I have nothing against the free expression of even the most fanciful eccentricities; but it is quite another matter to try and pass off worthless content as scientific and substantiated information (through the use of "Ph.D." and the data on the jacket). It is purely and simply fraud, and legal action should be taken against fraud of this nature. It is only to be hoped that no professional (police officer, judge, lawyer, etc.) whose decisions about a person's honesty affects their destiny reads this book or takes it seriously.

A dramatic example of the potential practical consequences of the dissemination of unscientific data or procedures concerns the controversial training programme by Inbau, Reid, Buckley and Jane (2001). Imparted by the company of *John E. Reid & Associates*, this training programme was designed for members of the police and security services who had to interrogate suspects. The company boasts of having trained more than 300,000 professionals since its first seminar on interrogations and interviews in 1974 (see http://www.reid.com). Part of the Inbau et al. (2001) programme focuses on deception cues. However, the cues it teaches are not the few shown by empirical research to be of possible use (see the interesting study by Blair & Kooi, 2004); also, paying attention to such cues reduces the accuracy of police in judging the credibility of true statements (Mann, Vrij & Bull, 2004). Moreover, Kassin and Fong (1999) have shown empirically that training in the Inbau et al. cues produces a *reduction* in the overall accuracy achieved, accompanied by a bias towards saying that subjects are lying and increased confidence in one's judgements.

If we take into account that, in many countries, before submitting the suspect to a strict interrogation the police carry out a more relaxed interview in order to establish innocence or guilt on the basis of behavioural cues of deception, the danger of the misinformation provided by *John E. Reid & Associates* becomes clear. But this danger is magnified if we consider the type of interrogation proposed by the *Reid & Associates* programme, involving as it does a highly aggressive and coercive approach that can lead many innocent people to confess to the crime being investigated (e.g., Kassin, 2005; Kassin & Gudjonsson, 2004). Briefly, the police: (a) interview the suspect; (b) observe certain behavioural cues of scarce diagnostic value, but which they believe to be associated with de-

[1] For example, Garrido, Masip & Herrero (2004) found that police are considered to be more capable of differentiating between truth and lies than the general population.

ception, and in consequence conclude that the suspect is lying; (c) on the basis of this conviction, submit the suspect to a process of tough interrogation –so tough that it leads many innocent suspects to confess (Kassin, 2004, 2005; Kassin & Gudjonsson, 2004). This process may explain a large part of the numerous cases reported in countries such as the United States (where Inbau and Reid's technique enjoys some popularity among members of the security services) of people who have been imprisoned on the basis of a confession that has later been reliably demonstrated to have been false (Drizin & Leo, 2004).

The aim of the present study is to "dismantle" a series of erroneous popular beliefs, in many cases spread through courses imparted or books written by people with little or no relevant qualifications, in relation to a clearly "psychological" topic, that of the detection of deception from non-verbal behaviour. The information presented on the following pages is based on the most rigorous scientific research in psychology and interpersonal communication. This information will be of undoubted interest for psychology professionals for three reasons: a) it represents a part of their discipline; b) the utility it may have in many areas of psychology; and c) in view of the consultant role of psychologists, who must respond to the call of other professionals and do so in accordance with psychological science, questioning the misleading beliefs the inquirer may hold.

ACCURACY: IS THE LIAR CAUGHT SOONER THAN THE CRIPPLE?

A widely held popular belief is that which is reflected in the saying "the liar is caught sooner than the cripple". In other words: it is easy to catch a liar. Is this belief correct?

Observers' accuracy (rate of hits) on making assessments of credibility (truth or lies) has been one of the aspects most widely studied in the field of deception. The experimental procedure employed usually consists in presenting a sample of observer or receiver subjects with a series of statements made by a group of emitter subjects (the potential liars). These statements are presented in audiovisual or auditory format, using tape recordings or "live" performances (see Chap. 3 of Miller & Stiff, 1993, for a description of the experimental paradigms employed). In some cases emitters and receivers are allowed to interact freely (Buller & Burgoon, 1996). Receivers must indicate, usually on a form, whether each statement is true or false. Sometimes they are also required to indicate their level of confidence in their judgement and the cues that led them to make their decision.

Normally, half of the statements presented are true and the other half are false. Thus, by chance alone, the observers can get half of their judgements right –that is, they can obtain an accuracy of 50%. What is the accuracy actually achieved in empirical studies? In 1980, Kraut published a review of the studies carried out up to that year, which indicated a mean accuracy of 57%. Twenty years later, Vrij (2000) calculated the average rate for 39 relevant studies. The result was almost identical to that of Kraut's review: 56.6%. Approximately one third (n = 12) of the experiments reviewed by Vrij showed an accuracy situated in the narrow range of 54% to 56%. In no experiment was the accuracy below 30% or above 64% (Vrij, 2000).

More recently, much more exhaustive and up-to-date reviews have been carried out, based on more meticulous sampling of the studies. Aamodt and Mitchell (in press) performed a meta-analysis on the effect of various individual variables on the accuracy of credibility judgements. Examining a total of 193 different samples of receivers, with a total number of 14,379 observers, they obtained a mean accuracy of 54.5%. In a more extensive study (including a total of 349 samples of receivers, with 22,282 subjects who assessed the credibility of messages from 3864 emitters), Bond and DePaulo (in press) found a mean accuracy of 53.4%. Even though this is significantly higher than the 50% expected by chance, in absolute terms it is an extremely poor accuracy rate. It means that of every 100 messages, 47 are judged erroneously. That is, we have almost the same probability of getting our judgements right as we have of getting them wrong. The accuracy of human detectors in judging credibility on the basis of observing behaviour is, despite the claims of popular wisdom, extremely limited. Indeed, of the different approaches to the detection of deception, the non-verbal one is that which gives the lowest levels of accuracy[2].

This limitation extends, likewise, to those professionals for whom lie detection is important, and who have experience in tasks of assessing credibility. Thus, compared to the 54.2% obtained by lay university students, Aamodt and Mitchell (in press) report levels of 50.8% for samples of detectives, of 54.5% for American federal agents, of 55.3% for police and customs officials, of 59.0% for judges, and of 61.6% for the four samples of psycholo-

gists included in their meta-analysis. Bond and DePaulo (in press) use contrast statistics for comparing the accuracy of "experts" (security service personnel, judges, psychiatrists, auditors, etc.) and "non-experts". Neither in the intra-study comparisons (on considering jointly all the experiments in which this comparison had been made) nor in the inter-study comparisons (comparison of the accuracy level in experiments in which the observers had been "experts" with those in which they had been "non-experts") were the differences found to be significant. In the inter-study comparisons the accuracy levels obtained were 52.9% for the "experts" and 56.9% for the "non-experts". In sum, professionals familiar with deception are *not* better detectors than lay observers.

Not only is accuracy low, but, moreover, it is *uniformly* low. There is evidence of a set of situational and personal factors that influence judgements and accuracy levels in a statistically significant way (Masip, Garrido & Herrero, 2002b). Thus, Bond and DePaulo (in press) found that certain variables (communication channel, emitter's motivation, preparation, previous exposure to emitter's behaviour and emitter-receiver interaction vs. non-interaction) had a significant impact on the rate of hits[3]. However, it is true that for some of these (motivation and preparation) this impact only appeared in the intra-study comparisons, and not in the inter-study ones. Furthermore, despite the significance of some differences, practically in all cases in which the authors report accuracy rates they were below 60%. Thus, the influence of these variables, despite its statistical significance, is really quite low in absolute terms. In the meta-analytical study by Aamodt and Mitchell (2005), the authors show that such important individual variables as receivers' age, sex, educational level/cognitive ability and traits of extraversion and neuroticism are not significantly related to accuracy of judgements. Only self-monitoring appears to show a weak positive relationship with it (r = .14).

These results refer to the detection of lies and truth (they reflect the percentage of correct classifications on considering true and false statements jointly), but what specifically occurs in the case of the detection *of lies*? Research shows that people more easily identify truth than lies (Levine, Park & McCornack, 1999). This is because we have a tendency to consider that others are telling the truth, which increases our accuracy on judging truths and reduces it on judging lies (Levine et al., 1999; Masip et al., 2002b). Thus, for example, the meta-analysis by Bond and DePaulo (in press) found the mean percentage of truth judgements was 55.0%, significantly higher than the 50% expected by chance. This meant that accuracy on judging true statements was 60.3%, markedly higher than that for judging false statements, for which the rate was just 48.7%.

This tendency to judge statements as true may be due to a variety of factors (see Levine *et al.*, 1999). It may be

[2] A recent official report of the *British Psychological Society* by Bull, Baron, Gudjonsson, Hampson, Rippon and Vrij (2004) presents the results of various reviews on the validity of the polygraph. Using the Control Question Test (CQT), the percentage of liars identified ranges, depending on the review considered, from 83% to 89%, and the percentage of truth-tellers identified ranges from 53% to 78%. Using the Guilty Knowledge Test (GKT), the polygraph permits identification of practically all the truth-tellers (accuracy of 98% and 94%, depending on the review considered), but shows poor capacity for detecting liars (42% and 76%) (Bull et al., 2004). Notable among the verbal procedures for assessing credibility are Criteria-Based Credibility Assessment (CBCA) and Reality Monitoring (RM). CBCA permits correct identification of 73% of true statements and 72% of false statements (Vrij, 2005). Accuracy of RM is similar, attaining a discrimination level of 72% for the classification of both true and false statements (Masip, Sporer, Garrido & Herrero, 2005). As we pointed out elsewhere (Masip, Garrido & Herrero, 2002b), in contrast to the polygraph users or assessors who employ CBCA and RM, the observers in the experiments carried out from the non-verbal approach are not trained, so that comparison is inappropriate. Nevertheless, as pointed out later in this article, the increases obtained through training in non-verbal indicators are quite limited. A methodology that produced good results from the analysis of non-verbal behaviour is that employed by Vrij, Edward, Roberts and Bull (2000), even if their findings need to be replicated. On this issue, see Masip et al. (2002b).

[3] More precisely, accuracy was lower when observers were exposed to the visual channel than when they were exposed to the auditory and audiovisual channels; the intra-study comparisons (but not the inter-study ones) showed that it is easier to detect motivated emitters than non-motivated emitters; also only in the intra-study comparisons, accuracy was lower when emitters had been able to prepare the message than when they had not prepared it; previous exposure to the emitter's habitual behaviour favoured detection; and the intra-study comparisons (inter-study comparisons could not be made because this factor varied only on a few occasions) indicated that detection is greater when there is emitter-receiver interaction than when the receiver observes a continuous and uninterrupted message from the emitter (Bond & DePaulo, in press).

based on a heuristic form of processing (Stiff, Kim & Ramesh, 1992), or on the actual functioning of the mind, which would in principle represent as true all the incoming information it understands (Gilbert, Krull & Malone, 1990); alternatively, the tendency may derive from the adaptive strategy of believing the messages received, since in everyday life the majority of them are true (Anderson, Ansfield & DePaulo, 1999). Recently, on the basis of two studies showing that the higher the quantity of information provided to the receiver, the less marked the bias towards truth, we have proposed that this bias may be due to an experimental artefact (Masip, Garrido & Herrero, 2005, in press). Certainly, in the research carried out to date, the fragments of behaviour of the emitter used as stimulus material have been very small, and this has limited the quantity of information observers can receive from the emitters, so that, on forming their judgements, observers are obliged to use a heuristic form of processing. And in credibility assessment tasks, heuristic judgements tend to be truth judgements (see Gilbert et al., 1990; Millar & Millar, 1997; Stiff et al., 1992). Thus, the truth bias found in research may be due to the brevity of the behavioural samples employed. In line with this idea, we have shown that the use of more extensive and informative samples of behaviour reduces this bias (Masip, Garrido & Herrero, 2005, in press). Nevertheless, this finding needs to be replicated by other research teams, and there are still some unanswered questions in relation to it (Masip, Garrido & Herrero, 2005, in press).

In any case, the tendency to judge statements as true appears to be weaker among those professionals for whom lie detection is more relevant than among others (Bond & DePaulo, in press). It has even been claimed, on the basis of empirical results, that such professionals actually present an opposite bias that leads them to judge statements as false (Meissner & Kassin, 2002), and that they have a general tendency to question the truth of what others say[4] (Masip, Alonso, Garrido & Antón, 2005).

In sum, the research reviewed in this section shows that: (a) the capacity of human beings to discriminate between true and false messages is quite poor; (b) this is the case even among people for whom such discrimination has professional importance; (c) although there are some variables that significantly affect hit rates, in absolute terms the range of variation is from 50% to 60%, always remaining below acceptable levels of accuracy; (d) research shows that we tend to believe what others say to us, so that we detect more truths than lies; however, there is evidence to suggest that this result may be due to the way in which the research has normally been carried out; and (e) on the other hand, professionals for whom the assessment of credibility is important display a tendency to consider messages as *false*.

CONFIDENCE: ARE WE AWARE OF OUR (IN)ABILITY TO DETECT LIES?

Having established the difficulty of detecting lies on the basis of non-verbal behaviour, we move onto another question examined by research: is there any relationship between the confidence we place in our judgements and our accuracy? DePaulo, Charlton, Cooper, Lindsay & Muhlenbruck (1997) carried out a meta-analysis of research on confidence about credibility judgements. With the sample of 18 relevant studies they were able to locate, they found a mean correlation that was practically null: r = .04. Aamodt and Mitchell (in press) examined the same issue, adding more recent experiments to those included in the DePaulo et al. (1997) meta-analysis. The mean correlation in 58 studies found by Aamodt and Mitchell is virtually the same: r = .05. In sum, it seems that people are unaware of the correctness or incorrectness of their credibility judgements.

Another interesting finding related to confidence concerns the evidence that we tend to overestimate our ability to discriminate between truth and lies. DePaulo et al. (1997) compared confidence and accuracy in six studies in which both variables had been measured on a scale of 0 to 100 (or whose scores could be transformed into these values). They found a mean accuracy of 57.20% and mean confidence in judgements of 72.91%, clearly superior.

CUES: LOOK ME IN THE EYE AND TELL ME THE TRUTH

Many popular books on non-verbal communication present lie detection as a simple task: all we need to do is observe whether emitters display certain clearly visible behavioural signals to determine whether they are lying

[4] Recently, Kassin, Meissner and Norwick (2005) found that police tend more than non-police to consider as *true* a series of *false* confessions of crimes. This has led these authors to modify their initial view and to maintain that, more than a bias towards considering statements to be false, what such professionals present is a bias towards considering that the emitters of such statements are guilty.

or not. For example, Lieberman (1998) and Pease (1981/1988) claim that covering the mouth, touching the nose, rubbing an eye or the neck or pulling one's shirt collar are signs that a person is lying.

Likewise, people have very clear beliefs about what the behavioural cues of deception are (see, among other reviews, those of Strömwall et al., 2004 or Vrij, 2000). For example, an extremely widespread belief (and which is also found in Lieberman's book) is that liars avert their eyes. In a recent transcultural study this stereotype was found to have universal validity. When people from 58 countries were asked "How can you tell if someone is lying?", those from 51 mentioned that people avert their eyes when they lie (Global Deception Research Team, in press). A second study used a questionnaire with closed questions, one of which referred to eye contact. The three response options were that people look you in the eye *more* when they are lying than when they are telling the truth, that they look *less*, and that they look *to the same extent*. In 61 of the 63 countries studied the participants chose the second of these three options more frequently than either of the other two (Global Deception Research Team, in press). How far are these beliefs correct? Are there clear indicators of deception? What are they?

Various reviews have compared the results of studies focused on *real* deception cues (behaviours that differentiate true and false accounts) with those of studies examining *perceived* cues or people's *beliefs* about indicators of deceit and lying. Perceived cues are those that people *actually use* for making their credibility judgements, and beliefs are the cues that people *say* are useful for discriminating between truth and lies[5] (Masip & Garrido, 2000, 2001). In general, coincidences between these last two categories and the first are extremely scarce, reflecting the fact that people are largely unaware of the cues that can actually discriminate between true and false communication (Burgoon, Buller & Woodall, 1994; DePaulo, Stone & Lassiter, 1985; Vrij, 2000). For example, Vrij (2000) observed that, while people believe that, compared to truth-tellers, liars move their extremities more, avert their eyes more, blink more, smile more, fidget and gesture more, change their position more and move their trunk more, the results of empirical research show that, in fact, liars move their extremities *less* than truth-tellers, and that the relationship between the rest of the behaviours and deception is not significant. Other popular beliefs examined by Vrij, such as that liars make more errors and hesitate more in their speech, make more pauses, and so on, have not received clear support from research, which has produced contradictory results due to the fact that certain variables, such as the cognitive complexity of the lie, may mediate the expression of relevant behaviours. There are two popular beliefs which, according to Vrij, are correct: that when people lie they speak in a slightly higher tone of voice, and that pauses are longer when people are lying than when they are telling the truth. In conclusion, then, the overwhelming majority of popular beliefs about non-verbal indicators of deception are erroneous. Unfortunately, the same applies to the beliefs of professionals such as police, judges, etc., which overlap to a large extent with those of the average citizen (see Strömwall et al., 2004, for a fuller discussion).

A possible explanation for this lack of agreement between beliefs and reality is provided by Kelley (1992), who hypothesizes that common sense notions are probably less valid when they refer to the *microlevel* than when they refer to the *mesolevel*. At the microlevel, Kelley situates "events that occur rapidly ..., on small scales of magnitude or mass (e.g., small contractions of the facial muscles or changes in direction of gaze), and often invisibly..." (Kelley, 1992, p. 6). The mesolevel is the "level of molar individual behaviour..." (Kelley, 1992, p. 6), and includes "immediate and direct consequences, periods of time from minutes to days... This level is the centre of attention in everyday life..." (Kelley, 1992, p. 6). Without doubt, the identification of discrete cues of deception belongs to Kelley's microlevel.

Whatever the case, the discrepancy between popular stereotypes and empirical reality may explain the low

[5] *Real* deception cues are studied by comparing the extent to which various behavioural categories (e.g., direction of gaze, stammering) are present in true and false messages. In order to examine *perceived* deception cues a comparison is made between messages *judged* as true and messages *judged* as false by observers. *Beliefs* or *stereotypes* about deception cues are studied by asking people which cues they think can differentiate true accounts from false ones. As we saw on presenting the results of the work of the Global Deception Research Team (in press), open or closed questions can be used. Moreover, these can be formulated in general terms ("How can you tell if someone is lying?") or, as is the case in Masip, Garrido, Herrero, Antón and Alonso (in press), they can refer to a specific judgement or set of judgements ("On what did you base your conclusion that this person was lying/telling the truth?").

value of behavioural cues for formulating *correct* judgements about lying. Park, Levine, McCornack, Morrison and Ferrara (2002) asked a group of students to recall a case in which they had discovered that another person had lied to them and to indicate which strategies that had used on that occasion to discover the deception. The results show that the methods most commonly used were information from third persons, material evidence and the confession from the liar him/herself. The consideration of non-verbal and verbal cues was among the strategies least employed (2.1%). In sum, the role of such cues in formulating correct judgements about lying is minimal[6].

The work by Vrij (2000) described above reviews only a part of the literature. Subsequently, DePaulo, Lindsay, Malone, Muhlenbruck, Charlton and Cooper (2003) published the most exhaustive meta-analytical work carried out to date on non-verbal and verbal deception cues. Although they do not compare these indicators with popular beliefs, their results are extremely interesting, since they permit the isolation of cues which are of potential utility for discriminating between truth and lies. DePaulo et al. examined a total of 116 research reports that explore the relationship between 158 behavioural cues and the act of lying or telling the truth. The authors distinguished between two sets of cues. First, those that had been examined on at least three different occasions, for at least two of which the effect size had been able to be calculated. The effect size is, in this case, an index of the relationship between the presence/absence of the cue and whether the emitter is lying or telling the truth. It can only be calculated accurately if sufficient information is provided in the original research reports, which was not the case in all those examined by DePaulo et al. The second set of cues included all the rest. The calculations referring to the first set are more valid, given the larger number of samples and the greater accuracy in the calculations of effect size.

The authors found that just 24 cues of the 88 in the first group distinguished between true and false statements. Added to these were 17 from the second group. Overall, 24 + 17 = 41 cues from a total of 158 examined: just 26.0%. If we consider only the 24 significant cues from the first group, whose calculation gave more guarantees, the percentage would be 15.2%. In conclusion, and in contrast to what is argued in a series of "self-help" books and the claims of popular wisdom, there are very few differences between people's behaviour when they lie and when they tell the truth.

With the aim of isolating the most valid deception cues, DePaulo et al. (2003) concentrated on those based on a number of comparisons higher than five and with an effect size of 0.20 or more in absolute values. They found only 12 such cues, the majority of a verbal nature. The most discriminative cue (d = -0.55) seems to be verbal and vocal immediacy. This means that when they lie people respond in a less direct, clear and relevant way than when they tell the truth, and that they also do so in an evasive and impersonal manner (DePaulo et al., 2003). Moreover, by comparison with the accounts of people who are telling the truth, those of liars will appear more ambivalent and discrepant (e.g., there will be a lack of agreement between what is expressed through some channels and through others) (d = 0.34). Likewise, lies will have more details, (d = -0.30), a less logical structure (d = -0.25) and less contextual elaboration (d = -0.21) than truths. These are three verbal criteria of Criteria-Based Credibility Assessment, or CBCA[7] (Garrido & Masip, 2000, 2004; Masip, Garrido & Herrero, 2003; Vrij, 2005). False accounts will also appear more plausible (d = -0.23) and will contain more negative assertions and complaints (d = 0.21) than true ones. The speaker will appear insecure and hesitant, reflected in both the voice and the words (d = 0.30), will give the impression of being more nervous or tense (d = 0.27), will have a tense-sounding voice (d = 0.26) and a higher tone of voice (voice frequency) (d = 0.21). Furthermore, the personal involvement of the speaker at a verbal and non-verbal level will be lower in false accounts than in true ones (d = -0.21). It is important to point out that none of the colourful cues described by Pease

[6] Park et al. (2002) interpret their results as indicating that people do not employ verbal and non-verbal cues to make their judgements of credibility. However, since the authors confined themselves to examining lies that were discovered, we can only conclude that such cues have a limited effect on correct judgements about lying. It may be that these cues are frequently used but have very little discriminative power.

[7] The *logical structure* implies that the different details describe an identical course of events, the statement as a whole is coherent and logical and its parts "fit together". By *contextual elaboration* we understand that the event described is situated within a rich and complex spatio-temporal context (see Garrido & Masip, 2001).

(1981/1988) are on the list based on rigorous meta-analysis of the relevant research, and that nor is eye contact[8].

It is extremely important to bear in mind that these results are based on the *whole* set of studies and experimental conditions of the works analyzed by DePaulo et al. (2003). But a series of circumstances were detected that influence the utility of cues for discriminating between true and false statements. Thus, the emitter's motivation, the aim pursued with the deception (concealing a transgression vs. other purposes), the length of response (time during which emitters express themselves) and previous preparation of the lie influence the meaning and discriminative power of various cues (DePaulo *et al.*, 2003; DePaulo & Morris, 2004). For example, when the account was not prepared in advance, the response latency (time elapsed between the end of the question and the beginning of emitter's response) was greater for lies than for truth-telling, but when the account was prepared in advance the latency was greater for *truth-telling* than for lying. Similarly, there were various cues (e.g., blinking) that discriminated when emitters were lying about transgressions but did not discriminate when they were lying about other things (for a full description of the effects of the moderating variables on the cues, see DePaulo et al., 2003; DePaulo & Morris, 2004). In sum: (a) the meaning of the same cues (e.g., response latency) may change according to the circumstances; (b) there are behaviours (e.g., blinking) that discriminate significantly in some circumstances but not in others; and (c) there are cues (e.g., blinking) that do not discriminate in general terms but do so in highly specific circumstances, and vice versa. Thus, in contrast to the claims of many "self-help" books, not only are there few deception cues, but these are also highly specific to each situation. As Kelley (1992) points out, common sense is more sensitive to the principal effects than to the interactions revealed by science, and moreover, science discovers underlying factors not perceived by the lay observer, and which nevertheless strongly influence the results.

TRAINING: IS THERE ANY REMOTE HOPE?

The picture emerging so far is by no means an encouraging one: human beings are terrible lie detectors, our confidence levels are not related to the accuracy of our judgements, we tend to overestimate our ability to detect lies, our beliefs about deception cues are erroneous and we use the wrong cues on making our judgements. Is there any hope of our learning to do it correctly?

Numerous attempts have been made to train people to detect deception (see the reviews by Bull, 2004; Frank & Feeley, 2003; or Vrij, 2000). Vrij observes that three types of training have been employed. One consists in providing subjects with *feedback* on their results, so that they can learn from their errors and correct judgements as they make their credibility assessments. Another type of training is based on an *informational strategy*, consisting in indicating to observers the true relationship between certain cues and deception. A third type of training is based on an *attentional strategy*, whereby observers' attention is focused on certain revealing cues (without necessarily explaining their meaning), or on the most transparent channels (e.g., the auditory channel). According to Vrij, regardless of the method used, observers have in general managed to increase their level of hits in the training condition. However, the author also notes that such increases have been quite poor: a mean accuracy of 54% in the non-trained groups vs. 57% in the trained groups.

In a later work than that of Vrij (2000), and a more systematic one, Frank and Feeley (2003) meta-analyzed the research carried out to date on non-verbal training in lie detection. Their study considers 20 comparisons made in 11 published works, with a total of 1072 observers in the training groups and 1161 in the control groups. They found the increase in accuracy due to training to be statistically significant, but very small: they report a mean hit rate of 54% in the non-trained groups and of 58% in the trained groups; note that the values are almost identical to those found by Vrij (2000). These authors argue that poor quality of the training programmes employed may be behind such a small increase. However, while it

[8] The effect size for eye contact was $d = 0.01$, and for averting the eyes, $d = 0.02$; both ds were non-significant. The cues that gave effect sizes larger than 0.20 in absolute values but that were calculated on the basis of 5 or more comparisons (in reality, 3 to 5 comparisons) were cooperativeness ($d = -0.66$), admission of lack of memory ($d = -0.42$), dilation of pupils ($d = 0.39$), duration of account ($d = -0.35$), related external associations ($d = 0.35$), verbal immediacy ($d = -0.31$), spontaneous corrections ($d = -0.29$), raising of the chin ($d = 0.25$), attributions about the mental state of the other person ($d = 0.22$), repetitions of words and phrases ($d = 0.21$) and self-disapproval ($d = 0.21$). Positive values of d indicate that the behaviour is presented more on lying than on telling the truth; negative values have the opposite meaning.

is true that the programmes used present a series of limitations, a more fundamental problem concerns the scarce relationship, mentioned previously, between behavioural cues and deception, as well as the dependence of this relationship on diverse circumstances (DePaulo et al., 2004). This may have a negative impact on the effectiveness of the three forms of training identified by Vrij (2000). Thus, what can be learned through *feedback* in a programme of the first type will be confused, relative and of little value. In the case of an *informational strategy*, there will be little consistent and valid information at a transituational level that can be provided to observers. Finally, the use of an *attentional strategy* also presents problems. If observers are guided to focus their attention on certain *discrete cues*, these will necessarily have limited validity, and dependent on circumstances. And if the aim is to focus observers' attention on the auditory and audiovisual *channels*, significantly more transparent than the merely visual channel in the meta-analysis by Bond and DePaulo (in press), it should previously be borne in mind that, in the inter-study comparisons (Bond & DePaulo do not present the specific accuracy indices in the intra-study comparisons), the average levels of accuracy attained for such channels were 53.7% (auditory channel) and 53.9% (audiovisual channel), as against 50.2% for the visual channel. Remember that the hit rate by chance is 50%, and that total accuracy corresponds to 100%. Consequently, the final accuracy that can be achieved by observers will be quite low if we ask them to pay attention to the auditory or audiovisual channels.

Based on a partial analysis of the relevant research, Meissner & Kassin (2002) suggest that, rather than increasing accuracy, what training programmes do is increase observers' tendency to say that messages are false. In line with these appreciations, in the more extensive meta-analysis by Frank and Feeley (2003), the increase due to training was null on judging truth (accuracy of 58% in the non-trained groups vs. 56% in the trained groups), but substantial on judging lies (49% vs. 55%). This effect should come as no surprise. Although Vrij (2000) identified the three approaches described above, in reality, the majority of training programmes have been based on the strategy of informing observers about the supposed relationship between certain behavioural cues and deception. Normally, such training focuses specifically on the indicators of *lying*, and not on the indicators of truth. Certain behaviours are pointed out, trainees are told that these tend to appear more frequently when people are lying than when they are telling the truth, and they are invited to try and identify them in the experimental videos to determine whether emitters *are lying* (and not to decide whether they are *lying or telling the truth*). But the fact that certain cues appear *more frequently* in liars than in truth-tellers does not mean that they appear *exclusively* when people are lying. Thus, observers actively seek these deception cues, and as soon as they perceive the slightest hint of them, come to the firm decision that the emitter is lying. This may be why the training programme increases only the frequency of lying judgments, and not the accuracy on judging truths. Quite probably, a training programme focused on *truth* cues, or indeed a more balanced one that presented, with identical emphasis, indicators of *truth* and of *lies* (their opposites), and in which the task did not consist in *detecting lies*, but rather in *discriminating between true and false statements*, would have quite different effects. Our most recent research is exploring this possibility.

CONCLUSIONS

Popular wisdom maintains that "the liar is caught sooner than the cripple". The majority of people show great confidence in their assessments of truth and lies. There are, moreover, clear popular stereotypes about people's behaviour when they are lying. Likewise, bookshops and libraries abound with "self-help" books, widely read and accepted, which present lie detection on the basis of non-verbal behaviour as a simple task to learn, and which provide long lists of supposed indicators of deception with universal validity.

As a counter to popular beliefs and the claims of the so-called "self-help" books, the present article has discussed the results of several decades of rigorous research carried out by psychologists and communicologists. It is important for the reader to bear in mind that the majority of the findings described here come from wide-ranging meta-analytic studies, so that the samples are extremely large and heterogeneous (and hence, representative), and the results faithfully reflect the global findings of virtually all the research carried out to date. These results are in stark contrast to the suggestions of popular beliefs and the proposals of most "self-help" books. Thus, the following conclusions can be drawn: (a) the capacity of the human being for discriminating between truth and lies is extremely limited; this is the case even for professional groups for whom the detection of deception is an

important part of their work; (b) people are unaware of the correctness or incorrectness of their credibility judgements; (c) we tend to overestimate our ability to identify truth and lies; (d) we use the wrong cues on making credibility judgements; (e) popular beliefs about deception cues are mistaken; (f) the beliefs of professionals for whom the detection of deception is an important part of their work are also erroneous, and similar to those of other people; (g) it is *not* been demonstrated that the behavioural cues mentioned in the majority of "self-help" books permit adequate discrimination between truth and lies; (h) there are very few behaviours that truly permit us to distinguish between truth and lies; (i) in contrast to what we are led to understand by many "self-help" books, and what popular wisdom maintains, the meaning and discriminative power of behavioural cues depend on a series of situational variables; (j) also in contrast to the assertions of certain books addressed to the general public, learning to discriminate between truth and lies is extremely difficult, as shown by the limited effectiveness of various training programmes; and (k) rather than raising overall accuracy, the training programmes in common use increase the bias towards saying that statements are false.

Sometimes, certain professionals whose work involves the assessment of credibility allow themselves to be led by their naive beliefs. In other cases, in a laudable effort to learn and to extend their professional skills, they seek information in books apparently written by reputable psychology professionals, but which are in fact the work of scarcely qualified authors who offer only spurious advice of no scientific worth whatsoever. Some go even further, and attend courses or seminars, but these are often imparted by people from outside the fields of psychology and communication, or by more experienced colleagues who, frequently with the best of intentions, confine themselves to passing on their commonsense intuitions and beliefs, out of touch with scientific progress in the relevant field of knowledge. In certain contexts, the consequences of a wrong credibility judgement can be devastating (the conviction of any innocent person; restriction of access to a given job, or its loss: and so on), hence the need for those making such judgements to receive the most rigorous and up-to-date information in the field of the detection of deception. Psychologists are among such people, but they have the added responsibility of acting as consultants for other professionals (and laypersons) about the true relationship between behavioural cues and deception. In this regard, I would like to have been able to offer a clear list of specific behavioural cues, clearly perceptible and unambiguous, as unquestionable indicators of lying. This is what the "self-help" books do, but, unfortunately, the reality is much more complex. That is indeed the lesson to be learned.

ACKNOWLEDGEMENTS

The author would like to express his thanks to Eugenio Garrido, Nuria Hernández and Roberto Vivero for kindly agreeing to read earlier versions of this work and making very helpful comments and suggestions.

REFERENCES

Aamodt, M. & Mitchell, H. (in press). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner*.

Abozzi, P. (1997). *La interpretación de los gestos*. Barcelona: Martínez Roca. (Originally published in Italian in 1996 by L'Airone Editrice, Rome, Italy).

Anderson, D. E., Ansfield, M. E. & DePaulo, B. M. (1999). Love's best habit. Deception in the context of relationships. In P. Philippot, R. S. Feldman & E. J. Coats (Eds.), *The social context of nonverbal behaviour* (pp. 372-409). New York: Cambridge University Press.

Blair, J. P. & Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science and Management, 6*, 77-83.

Bond, C. F., Jr. & DePaulo, B. M. (in press). Accuracy of deception judgments. *Personality and Social Psychology Review*.

Bull, R. (2004). Training to detect deception from behavioural cues: Attempts and problems. In P.-A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 251-268). Cambridge: Cambridge University Press.

Bull, R., Baron, H., Gudjonsson, G., Hampson, S., Rippon, G. & Vrij, A. (2004). *A review of the current scientific status and fields of application of polygraphic deception detection*. London: British Psychological Society.

Buller, D. B. & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*, 203-242.

Burgoon, J. K., Buller, D. B. & Woodall, W. G. (1994). *Nonverbal communication*. Columbus, OH: Greyden Press.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J. & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*(4), 346-357.

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.

DePaulo, B. M. & Morris, W. (2004). Cues to deception and indirect lie detection. In P.-A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 15-40). Cambridge: Cambridge University Press.

DePaulo, B. M. & Pfeiffer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*, 249-267.

DePaulo, B. M., Stone, J. I. & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323-370). New York: McGraw-Hill.

Drizin, S. A. & Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review, 82*, 891-1004.

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.

Garrido, E., Herrero, C. & Masip, J. (2004). Psicología jurídica y sentido común: Construcción social. *Revista de Psicología General y Aplicada*, 57, 395-406.

Garrido, E. & Masip, J. (2004). La evaluación del abuso sexual infantil. In las *Actas del I Congreso de Psicología Jurídica y Forense en Red* (CD-Rom). Madrid: Colegio Oficial de Psicólogos de Madrid.

Garrido, E., Masip, J. & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology, 39*, 254-275.

Gilbert, D. T., Krull, D. S. & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*(4), 601-613.

Global Deception Research Team (in press). A world of lies. *Journal of Cross-Cultural Psychology*.

Inbau, F. E., Reid, J. E., Buckley, J. P. & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersberg, MD: Aspen.

Kassin, S. M. (2004). True or false: "I'd know a false confession if I saw one". In P.-A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 172-194). Cambridge: Cambridge University Press.

Kassin, S. M. (2005). On the psychology of confessions. *American Psychologist, 60*, 215-228.

Kassin, S. M. & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior, 23*, 499-516.

Kassin, S. M. & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest, 5*, 33-67.

Kassin, S. M., Meissner, C. A. & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior*, *29*, 211-227.

Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology, 43*, 1-23.

Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, *30*, 209-216.

Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., Bar, R., Pierro, A. & Mannetti, L. (2005). Says who? Epistemic authority effects in social judgment. *Advances in Experimental Social Psychology, 37*, 345-392.

Lazarsfield, P. F. (1949). The American Soldier –an expository review. *Public Opinion Quarterly, 13*, 377-404.

Levine, T. R., Park, H. S. & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs*, *66*, 125-144.

Lieberman, D. J. (1998). *Never be lied to again.* New York: St. Martin's Press.

Mann, S., Vrij, A. & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology, 89*, 137-149.

Masip, J., Alonso, H., Garrido, E. & Antón, C. (2005). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, 35(5), 1046-1066.

Masip, J. & Garrido, E. (2000). La evaluación de la credibilidad del testimonio en contextos judiciales a partir de indicadores conductuales. *Anuario de Psicología Jurídica, 10*, 93-131.

Masip, J. & Garrido, E. (2001). La evaluación psicológica de la credibilidad del testimonio. In F. Jiménez (Ed.), *Evaluación psicológica forense 1: Fuentes de información, abusos sexuales, testimonio, peligrosidad y reincidencia* (pp. 141-204). Salamanca: Amarú.

Masip, J., Garrido, E. & Herrero, C. (2002a). La detección de la mentira mediante la técnica SCAN. *Revista de Psicopatología Clínica, Legal y Forense, 2*, 39-62.

Masip, J., Garrido, E. & Herrero, C. (2002b). La detección del engaño sobre la base de sus correlatos conductuales: La precisión de los juicios. *Anuario de Psicología Jurídica*, *12*, 37-55.

Masip, J., Garrido, E. & Herrero, C. (2003). El Análisis de Contenido Basado en Criterios (CBCA). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica, 15*(1), 75-92.

Masip, J., Garrido, E. & Herrero, C. (2004). La detección de la mentira mediante la medida de la tensión en la voz: Una revisión crítica. *Estudios de Psicología, 25*, 13-30.

Masip, J., Garrido, E. & Herrero, C. (2005). *Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias*. Submitted for publication.

Masip, J., Garrido, E. & Herrero, C. (in press). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *International Journal of Psychology.*

Masip, J., Garrido, E., Herrero, C., Antón, C. & Alonso, H. (in press). Officers as lie detectors. Guilty before charged. In D. Chadee & J. Young (Eds.), *Current themes and theories in social psychology*. St. Augustine: The University of the West Indies Press.

Masip, J., Sporer, S. L., Garrido, E. & Herrero, C. (2005). The detection of deception with the Reality Monitoring approach: A review of the empirical evidence.

Psychology, Crime, & Law, 11(1), 99-122. Meissner, C. A. & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*, 469-480.

Millar, M. G. & Millar, K. (1997). The effects of cognitive capacity and suspicion on truth bias. *Communication Research, 24*(5), 556-570.

Miller, G. R. & Stiff, J. B. (1993). *Deceptive communication*. Newbury Park, CA: Sage.

Myers, D. G. (2000). *Psicología social* (6th ed.). Santa Fe de Bogotá: McGraw-Hill. (Originally published in English in 1999 by McGraw-Hill, New York, USA).

National Research Council. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press.

Park, H. S., Levine, T. R., McCornack, S., Morrison, K. & Ferrara, M. (2002). How people really detect lies. *Communication Monographs, 69*, 144-157.

Pease, A. (1988). *El lenguaje del cuerpo. Cómo leer el pensamiento de los otros a través de sus gestos*. Barcelona: Paidós. (Originally published in English in 1982 by Sheldon Press, London, UK).

Schlesinger, A., Jr. (1949). The statistical soldier. *Partisan Review, 16*, 852-856.

Stiff, J. B., Kim, H. J. & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research, 19*(3), 326-345.

Strömwall, L. A., Granhag, P.-A. & Hartwig, M. (2004). Practioners' beliefs about deception. In P.-A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229-250). Cambridge: Cambridge University Press.

Teigen, K. H. (1986). Old truths or fresh insights? A study of students' evaluations of proverbs. *British Journal of Social Psychology, 25*, 43-50.

Vrij, A. (2000). *Detecting lies and deceit. The psychology of lying and the implications for professional practice.* Chichester: Wiley.

Vrij, A. (2005). Criteria-Based Content Analysis. A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*(1), 3-41.

Vrij, A., Edward, K., Roberts, K. & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*(4), 239-263.

Papeles del Psicólogo, 2005. Vol. 26, pp. 92-98

CRITERIA-BASED CONTENT ANALYSIS (CBCA) IN STATEMENT CREDIBILITY ASSESSMENT

Verónica Godoy-Cervera and Lorenzo Higueras
University of Granada

The purpose of this work is, on the one hand, to describe Criteria-Based Content Analysis (CBCA), a procedure focusing originally on the assessment of children's testimony credibility, and on the other, to present some of the research on this topic. In recent years, however, there has been increasing interest in the application of this procedure to adults, so that another objective of the study was to discuss studies on the use of this statement credibility analysis technique with adults. Finally, we discuss some disadvantages of the technique and future lines of research in relation to its use.

Este trabajo está dirigido por un lado a la descripción del Análisis de Contenido Basado en Criterios (CBCA), procedimiento enfocado originalmente a la evaluación de la credibilidad del testimonio de niños, y por otro, a la presentación de algunas investigaciones realizadas en torno a este tema. En los últimos años, sin embargo, ha habido un creciente interés por generalizar la aplicación de este procedimiento a adultos, por lo que otro de nuestros objetivos será comentar los estudios encaminados a emplear esta prueba en la evaluación de la credibilidad de las declaraciones de adultos. Finalmente comentaremos algunos inconvenientes de esta técnica y futuras líneas de investigación.

It is not uncommon for police investigations to have access to no more than the testimonies of the victim and the accused as evidence in a crime. Given this situation, several researchers have worked on the development of systematic methods that help to distinguish honest testimonies from those that have been fabricated. Vrij (2000), and Vrij, Edward and Bull (2001) have classified these procedures in three groups. The first of these includes procedures for recording and analyzing the psychophysiological activity of the person who is lying; the second group concerns the examination of the witness's non-verbal behaviour (Vrij, Edward, Roberts & Bull, 2000); and the third, on which we shall concentrate here, refers to the study of the content of the witness's statement (Masip, Sporer, Garrido & Herrero, 2005; Ruby & Brigham, 1997).

DEVELOPMENT OF CRITERIA-BASED CONTENT ANALYSIS (CBCA)

Statement Validity Assessment (SVA) is the technique most widely employed for assessing the truthfulness of verbal statements (Vrij, 2000). The SVA was developed in Germany, based on the clinical experience of several psychologists. It was around 1950 that Udo Undeutsch first described SVA (see Undeutsch, 1989), which was subsequently modified, on the way to its current form, by Steller and Köhnken (1989) and Raskin and Esplin (1991). SVA was initially developed for assessing the verbal statements of minors who had been the victims of sexual abuse. However, in recent years it has been attempted to validate and generalize the application of this instrument for adults (Vrij et al., 2001; Vrij et al., 2000). Despite the fact that it is an instrument widely used in the forensic context as psychological evidence, it should be considered not as a *test* or standardized scale, but as semi-standardized method for assessing the credibility of statements (Steller, 1989). The development of SVA is based on what Steller (1989) has called the Undeutsch hypothesis. According to this hypothesis, a testimony based on a real experience differs in quality and content from a testimony based on an imagined event.

Criteria-Based Content Analysis (CBCA) is the core component of SVA, and this has led to its being the element most frequently studied by researchers (Ruby & Brigham, 1997); it also serves as the focus of the present work. SVA is made up of three mutually dependent components: a) a structured interview with the victim, b) CBCA, which assesses the content of the person's testimony, and c) the integration of CBCA with the information obtained through a set of questions called the *Validity*

Correspondence: Verónica Godoy Cervera. *Facultad de Psicología, Universidad de Granada 18071, Granada, Spain. E-mail: vgodoy@ugr.es*

Checklist, so that the information provided by content analysis of the statement is combined with other relevant information of the case and with information obtained from the exploration of the interview or interviews previously carried out (Horowitz, 1991).

The interview should precede application of the CBCA criteria. The basic objective is to obtain material on which to apply these criteria. For the interview to be appropriately conducted, it is important for the interviewer to be familiar with the content of the criteria. Likewise, the interviewer should try to obtain the greatest possible amount of data by using an interview designed to maximize the quantity of information provided by the witness and minimize any type of contamination generated by either the interviewer him/herself or any other adult (Raskin & Esplin, 1991).

CBCA is applied to the content of the testimony, and its purpose is to determine whether the quality and specific content of that testimony are indicative of an account generated from memory traces or of one that is the product of invention, fantasy or the influence of another person. Any analysis made using CBCA will be influenced by the interview characteristics and by what the subject has or has not experienced. Throughout this process, it is important that the interviewer takes into account the person's age, experience and cognitive ability level (Raskin & Esplin, 1991). One of the main limitations of CBCA concerns the fact that it is applied to situations in which witnesses have information on the basis of which they can invent a charge that incorporates some of the criteria. For example, a child who has been sexually abused on a previous occasion may provide a false testimony that appears convincing as it is based on the memory traces from that previous experience. This aspect should be borne in mind on reviewing the case by means of the *Validity Checklist* (Raskin & Esplin, 1991). The verbal content of the statement is analyzed through the application of a series of 19 criteria (see Table 1), organized in five broad categories, and with the purpose of differentiating between true statements and fabricated statements. The basic idea is that a true testimony contains a greater number of criteria (for a detailed description of these criteria, see Steller and Köhnken, 1989).

Analysis of the interview by means of the 19 CBCA criteria is carried out by giving numerical scores to each one of the criteria. Steller's (1989) proposal is to assign 2, 1 or 0 points depending on whether the criterion is strongly present, present or absent in the statement, while other authors (Lamb, Sternberg, Esplin, Hershkowitz, Orbach & Hovav, 1997) propose awarding 1 for present and 0 for absent. Neither the interview nor the results obtained from the criteria are completely valid until they have been put into context by means of the *Validity Checklist*, which is made up of four general categories of information (Steller & Köhnken, 1989):

a) *Psychological characteristics*. In this category it is important to assess the appropriateness of language and affect and susceptibility to suggestion.

b) *Interview characteristics*. The assessor should analyze the quality of the interview, rating the type of questions asked (suggestive, leading or coactive questions) and its overall appropriateness.

c) *Motivation for making false accusations*. The information in this category should help to rule out those aspects of a motivational nature that may be influencing the person to provide a false testimony. It

TABLE 1
CRITERIA-BASED CONTENT ANALYSIS (CBCA). MODIFIED FROM STELLER AND KÖHNKEN (1989)

General Characteristics
1. Logical structure
2. Unstructured production
3. Quantity of details
Specific Content
4. Contextual embedding
5. Description of interactions
6. Reproduction of conversation
7. Unexpected complications during the incident
Peculiarities of Content
8. Unusual details
9. Superfluous details
10. Accurately reported details misunderstood
11. Related external associations
12. Allusions to subjective mental state
13. Attribution of the accused's mental state
Motivation-Related Content
14. Spontaneous corrections
15. Admitting lack of memory
16. Raising doubts about one's own testimony
17. Self-Deprecation
18. Pardoning the accused
Specific Elements of the Offence
19. Specific details of the offence

should also be borne in mind that a minor can be under pressure from a third person to make a false statement. An important aspect of this category is the assessment of the context in which the statement is generated.

d) *Aspects related to the investigation*. This section is designed with the aim of rating the consistency between previous statements and investigation results and medical reports.

The assessor must analyze the information from these four categories and, on the basis of this analysis, determine whether this information supports the person's testimony. Likewise, the objective of the *Validity Checklist* is to rate various explanatory hypotheses, examining all the available information on the case. Raskin and Esplin (1991) propose five hypotheses that should be considered by the assessor:

a) The statement is valid, but the child has substituted the identity of the aggressor by that of a different person.
b) The statement is valid, but the child has been influenced or has invented additional information that is not true.
c) The child has been put under pressure by a third person to formulate a false version of events.
d) Due to personal interests or to help third persons, the child has given a false statement.
e) As a consequence of psychological problems, the child has given a testimony based on fantasy or invention.

It is important to stress that the purpose of SVA is to assess the credibility of the content of statements, and not to assess the credibility of the persons themselves (Steller & Köhnken, 1989).

Another serious limitation of CBCA is the lack up to now of a decision rule that helps to establish how many criteria determine whether a statement should be classified as credible or not credible; even less consideration has been given to the weight each criterion should be assigned. Alonso-Quecuty (1999) proposes that the weight of each criterion be assigned on the basis of diverse factors, such as: number of previous interviews the child has had, complexity of the incident, age of the minor, and time elapsed since the event. Once the CBCA criteria and the *Validity Checklist* have been applied, the final result of the analysis permits the statement to be qualitatively classified according to five categories (Alonso-Quecuty, 1999; Steller, 1989):

- Credible.
- Probably credible.
- Indeterminate.
- Probably not credible.
- Not credible.

RESEARCH IN RELATION TO CBCA

Recent years have seen an increase in the number of studies on CBCA due to its extensive use in the forensic context. Studies have been of two basic types: 1) those that use real cases of minors allegedly the victims of sexual abuse, and in which other elements of the case are used as measures of truthfulness; and 2) experimental studies in which subjects are induced to manipulate their statement, providing either a true or false testimony (Ruby & Brigham, 1997).

Studies with children

Given the fact that CBCA was designed to be applied to the statements of minors alleged to be the victims of sexual abuse, the majority of published studies have employed samples with these characteristics. Let us briefly consider some of these studies carried out with children.

An important study, insofar as it focused on children presumed to have suffered sexual abuse, was that of Lamb et al. (1997). Their sample was made up of 98 Israeli children (28 boys and 70 girls) aged between 4 and 13 (mean 8.72). As measures of the truthfulness of the statement they employed other elements of the case, including material or physical evidence, the accused's testimony, and so on. As predicted, there was greater presence of the CBCA criteria in the credible accounts (mean 6.74) than in the not-credible accounts (mean 4.85). Nevertheless, the authors state that the differences found were not as significant as those of previous studies.

More recently, Santtila, Roppola, Runtti and Niem (2000) analyzed the effect of age, verbal ability (measured with the WISC-R verbal scale) and interviewer's emotional style on the presence of CBCA criteria in the statements of 68 children from three different age groups: 7-8, 10-11 and 13-14 years. In their experiment, they asked each child to give an account of two personal experiences, one real and one false. The results showed a correct classification rate of 66%. They also found that age and verbal ability of the minor, as well as interviewer behaviour, influenced the number of CBCA criteria present in the statements, regardless of whether

these were true or false. Likewise, they found that different criteria appeared in the statements depending on participants' age range. These authors suggest integrating the information from CBCA with that obtained through the *Validity Checklist*, and conclude that judicial decisions should not be based exclusively on the results provided by CBCA-SVA.

Research has also focused on analyzing the influence of the familiarity of the event to be recalled on the presence of CBCA criteria. For example, Pezdek et al. (2004) carried out an experiment with 114 children, hypothesizing that descriptions of familiar events were more likely to be classified as true than descriptions of unfamiliar events. The results suggested a greater presence of CBCA criteria in accounts of familiar events than in accounts of unfamiliar events.

In a similar line, Blandon-Gitlin, Pezdek, Rogers and Brodie (2005), with a sample of 94 children, analyzed the interaction between familiarity of the event and its veracity. They found the scores obtained through CBCA to be more strongly influenced by the familiarity of the event than by its truthfulness. In both this study and the previous one, the authors concluded by suggesting that CBCA, in its current form, is of limited utility as a tool for assessing statement credibility in minors.

Studies with adults

Given the good results obtained with children, it has been attempted to apply it to adults also. However, there are fewer studies than in the case of minors. Let us consider a few of them.

Some research has concentrated on identifying the criteria most commonly found in true statements, which would consequently be the most sensitive ones in the discrimination between true and false testimonies. For example, in a meta-analysis, Ruby and Brigham (1997) found the criteria most frequently found in true statements to be 1, 2, 3, 4, 5, 6, 7, 8, 12, 14 and 15. On the other hand, Köhnken, Schimossek, Aschermann and Höfer (1995) found that true testimonies included a larger number of details (Criterion 3) and were unstructured (Criterion 4), and that people had a greater tendency to admit lack of memory about the event (Criterion 15). They also found five of the six additional criteria included in their experiment to be significant, though while four of them (expression of insecurity, reporting style, justification of memory lapse and clichés) were significantly more likely in false statements, the fifth (repetitions), contrary to predictions, was significantly more frequently found in truthful accounts (for a detailed description of the additional criteria included in this study, see Köhnken et al., 1995). Likewise, Vrij, Akehurst, Soukara and Bull (2004a) reported that Criteria 1, 3, 4, 5 and 6 were the most effective for differentiating between true and false testimonies.

The vast majority of studies that analyze the effectiveness of CBCA have used samples of Europeans or other Caucasian people. Therefore, Ruby and Brigham (1998) decided to explore the differences that may arise between subjects from different ethnic groups. These authors started out from the notion that there are differences at a verbal level, in terms of style and content, between accounts provided by persons of different racial origin. The hypothesis they proposed was that, since CBCA was developed in a white European culture, its application would only be effective for discriminating between the testimonies of this type of subject, and that accounts by black subjects would include significantly fewer criteria. The results showed that CBCA functioned differently according to a person's race, and that different criteria were better predictors of truth for one ethnic group than for the other. In the true accounts given by the black subjects, the criteria that most frequently appeared were 3, 6, 12, 14, and 17, though a comparison of true accounts by white subjects with those from black subjects showed no criterion that appeared significantly more frequently. Considering the testimonies of the subjects of both races jointly, it was found that certain criteria appeared significantly more in true statements (Criteria 2, 5, 7, 8, 9, 14 and 15). However, in false accounts there was also greater presence of certain criteria: 1, 4, 11, 12 and 17. As regards the classification of the statements, the results were not conclusive. The researchers found that if they took as true all those statements in which 5 criteria were present, the percentage of correct classification for true statements was 89%; however, with this rule, a high percentage of false accounts were also classified as true (92%). On employing stricter decision criteria (taking into account the presence of 6 or 7 criteria), the number of true accounts correctly classified fell, and the number of false accounts correctly classified increased –that is, with this criterion fewer false accounts were classified as true.

Research has also revealed that there are differences in CBCA scores on comparing the results for children and adults. This is the case, for example, of the study by Vrij et al. (2004a). It has also been found that previous

knowledge of the content of the CBCA criteria has a negative influence on the validity of the instrument, and that subjects instructed to lie and who know the criteria in advance can provide statements that appear to be true (Vrij, Akehurst, Soukara & Bull, 2002).

The effectiveness of CBCA has also been compared with that of other procedures for assessing the content of statements, and researchers have explored whether the combination of these methods improves the classification of testimonies. Specifically, *reality monitoring* (Johnson & Raye, 1981) is the procedure with which it has most commonly been compared. *Reality monitoring* developed within the basic research context, and was initially applied in the clinical field (where it stimulated extensive research, e.g., Bentall, Baker & Havers, 1991; Brebion, Smith, Gorman & Amador, 1997; Harvey, 1985; Johnson, Raye, Hasher & Chromiak, 1979; Raye & Johnson,1980), before being employed in the forensic context (see, for a review, Mitchell & Johnson, 2000). It postulates that memories of what was actually seen have different characteristics from internally-generated "memories". The proposal of *reality monitoring* is that memories of external origin will have more contextual and sensory attributes, greater semantic detail, and less information on cognitive operations than memories of internal origin (Johnson & Raye, 1981).

One study that compares the results of CBCA and those of *reality monitoring* is that carried out by Sporer (1997). In his experiment, Sporer used a sample of 40 psychology students (20 men and 20 women), and participants were instructed to recount two personal experiences: one true and the other false. The author's objective was to explore the efficacy of CBCA and *reality monitoring* in the discrimination of fabricated and truthful accounts and to check whether the combined use of the two instruments improved the classification of the accounts. The results showed that CBCA was effective in 65% of total classifications, with 70% effectiveness in the classification of the true accounts and 60% effectiveness in that of the false accounts. As regards *reality monitoring*, 71.3% of the statements were correctly classified. Of the true statements, 75% were correctly classified, and of the false accounts, 67.5%. On combining CBCA and *reality monitoring*, the percentage of correct classification rose to 79%. In a later study, Vrij, Akehurst, Soukara and Bull (2004b) found that the true testimonies obtained higher scores both in CBCA and the criteria of *reality monitoring*, with classification effectiveness of 60% and 74%, respectively. Nevertheless, on integrating the results of the two instruments, no improvements were found, and the percentage of classification remained at 74%, so that the combination of the two techniques did not produce improvements on this occasion.

Various studies have also considered the option of combining procedures for assessing the verbal content of statements with behavioural indicators of deception. Vrij et al. (2001) found that those who lied scored lower in the CBCA and *reality monitoring* criteria and were more likely to present certain behaviours indicative of deception, such as waiting a long time before answering, talking more quickly, and so on. They even found that both CBCA and *reality monitoring* were the most sensitive tools in the detection of deceit in relation to other variables, such as non-verbal behaviour. In a previous study, Vrij et al. (2000) found that the joint use of verbal and non-verbal indicators of deception (CBCA and *reality monitoring*) led to an increase in the percentage of correct classifications. These results were ratified in a more recent study by Vrij et al. (2004a) that examined the verbal and non-verbal behaviour of children and adults.

CONCLUSIONS

Criteria-Based Content Analysis (CBCA) is still far from being a totally effective tool in the detection of deceitful testimonies, and requires a good deal of refining. Regardless of whether this technique is applied to children or adults, there are many factors that exert a negative influence and can affect its results. As shown by research, there are individual differences, such as age, verbal ability, interviewer's attitude (Santtila et al., 2000), familiarity of the event (Blandon-Gitlin et al., 2005; Pezdek et al., 2004), previous knowledge of the instrument, (Vrij et al., 2002) and ethnic group of the person (Ruby & Brigham, 1998), that have to be taken into account and controlled as far as possible when this technique is employed, and which, therefore prevent the immediate individual application of CBCA.

However, and although studies show that true statements contain a larger number of criteria than fabricated statements, the major disadvantage of CBCA is that there is no general consensus establishing a minimum number of criteria a statement should include in order to be classified as credible, or the weight that should be attributed to each of them. Landry and Brigham (1992) have proposed as a minimum the presence of five criteria for a statement to be classified as truthful. However, other au-

thors manipulated in their experiments the number of criteria in the classification of statements, and their results were not as conclusive as expected (Ruby & Brigham, 1998). Likewise, there is still a need to define the number of criteria that should be included in assessment of the testimonies of adults. Given that CBCA was developed for assessing the statements of minors, it is likely that some criteria do not work with adults, as is the case of Criterion 10 (Accurately reported details misunderstood). Thus, there is also a need for studies aimed at the definition of a group of criteria applicable to the testimonies of adults. As regards the weight each criterion should receive, in this aspect research is even further away from its objective, that of setting a general parameter. Nevertheless, certain criteria have been found to discriminate better than others between true and false testimonies (Ruby & Brigham, 1997, 1998). It may be that some of these criteria should receive higher scores in the general assessment of CBCA; however, it is essential to carry out further research with a view to clarifying this issue.

In accordance with the situation as presented here, a viable alternative in the detection of false testimonies is the combination of various techniques, such as *reality monitoring* and behavioural indicators of deception (Vrij et al., 2001; Vrij et al., 2000), which, as we have seen, on the majority of occasions improves the classification of statements.

Given the disadvantages mentioned, what does clearly emerge is that CBCA should be considered exclusively as a support tool, and never as the sole instrument on which to base judicial decisions (Santtila et al., 2000), at least for the present time.

REFERENCES

Alonso-Quecuty, M.L. (1999). Evaluación de la credibilidad de las declaraciones de menores víctimas de delitos contra la libertad sexual. *Papeles del Psicólogo, 73*, 36-40.

Bentall, R.P., Baker, G.A. & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Blandon-Gitlin, I., Pezdek, K., Rogers, M. & Brodie, L. (2005). Detecting deception in children: an experimental study of the effect of event familiarity on CBCA ratings. *Law and Human Behaviour, 29*, 187-197.

Brebion, G., Smith, M.J., Gorman, J.M. & Amador, X. (1997). Discrimination accuracy and decision biases in different types of reality monitoring in schizophrenia. *Journal of Nervous and Mental Disease, 185*, 247-253.

Harvey, P.D. (1985). Reality monitoring in mania and schizophrenia: the association of thought disorder and performance. *Journal of Nervous and Mental Disease, 173*, 67-73.

Horowitz, S.W. (1991). Empirical support for statement validity assessment. *Behavioural Assessment, 13*, 293-313.

Johnson, M.K. & Raye, C.L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, M.K., Raye, C.L., Hasher, L. & Chromiak, W. (1979). Are there developmental differences in reality monitoring? *Journal of Experimental Child Psychology, 27*, 120-128.

Köhnken, G., Schimossek, E., Aschermann, E. & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults` statements. *Journal of Applied Psychology, 80*, 671-684.

Lamb, M.E., Sternberg, K.J., Esplin, P.W., Hershkowitz, I., Orbach, Y. & Hovav, M. (1997). Criterion-based content Analysis: a field validation study. *Child, Abuse & Neglect, 21*, 255-264.

Landry, K.L. & Brigham, J.C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behaviour, 166*, 663-676.

Masip, J., Sporer, S.L., Garrido, E. & Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law, 11*, 99-122.

Mitchell, K. & Johnson, M. (2000). Source monitoring. Attributing mental experiences. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory*. New York: Oxford University Press.

Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G.S., Quas, J.A., Saywitz, K.J., Bidrose, S., Pipe, M.E., Rogers, M. & Brodie, L. (2004). Detecting deception in children: event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology, 89*, 119-126.

Raskin, D.C. & Esplin, P.W. (1991). Statement Validity Assessment: interview procedures and content analysis of children's statements of sexual abuse. *Behavioural Assessment, 13*, 265-291.

Raye, C.L. & Johnson, M.K. (1980). Reality monitoring vs. discriminating between external sources of memories. *Bulletin of the Psychonomic Society, 15*, 405-408.

Ruby, C.L. & Brigham, J.C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: a critical review. *Psychology, Public Policy, and Law, 3*, 705-737.

Ruby, C.L. & Brigham, J.C. (1998). Can criteria-based content analysis distinguish between true and false statements of African-American speakers? *Law and Human Behaviour, 22*, 369-388.

Santtila, P., Roppola, H., Runtti, M. & Niemi, P. (2000). Assessment of child witness statements using criteria-based content analysis (CBCA): the effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime & Law, 6*, 159-179.

Sporer, S.L. (1997). The less travelled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373-397.

Steller, M. (1989). Recent developments in statement analysis. In J.C. Yuille (Ed.), *Credibility assessment* (pp. 135-154). Dordrecht, Netherlands: Kluwer.

Steller, M. & Köhnken, G. (1989). Statement analysis: credibility assessment of children's testimonies in sexual abuse cases. In D.C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer.

Undeutsch, U. (1989). The development of statement reality analysis. In J.C. Yuille (Ed.), *Credibility assessment* (pp. 101-121). Dordrecht, Netherlands: Kluwer.

Vrij, A. (2000). *Detecting lies and deceit: the psychology of lying and the implications for professional practice*. Chichester: John Wiley & Sons.

Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behaviour, 26*, 261-283.

Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2004a). Detecting deceit via analysis o verbal and nonverbal behaviour in children's and adults. *Human Communication Research, 30*, 8-41.

Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2004b). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science, 36*, 113-126.

Vrij, A., Edward, K. & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulleting, 27*, 899-909.

Vrij, A., Edward, K., Roberts, K. & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behaviour. *Journal of Nonverbal Behaviour, 24*, 239-263.

Papeles del Psicólogo, 2005. Vol. 26, pp. 99-108

MALINGERING OF ORGANIC ILLNESS OR MENTAL DISORDER

Mercedes Inda Caro*, Serafín Lemos Giráldez*, Ana María López Rodrigo*
and José Luis Alonso Rionda**
** Psychology Faculty, University of Oviedo ** Mental Health Service of Asturias*

The purpose of this article is to analyze the nature and incidence of malingering and to describe several general measures used for determining symptom validity, as well as some instruments specifically created for assessing deception or malingering in clinical settings. In particularly, we consider some useful criteria for detecting the faking of post-traumatic stress disorder, organic brain syndromes, amnesia, and psychotic disorders.

Se analiza la naturaleza y la frecuencia de la simulación de síntomas clínicos, y se describen algunos instrumentos de medida generales, que han sido utilizados para determinar la validez de los síntomas, así como pruebas específicas para tomar decisiones respecto a la existencia de simulación en el ámbito clínico. En particular, se presentan algunos criterios de utilidad para evaluar la existencia de simulación de un trastorno de estrés post-traumático, de un síndrome orgánico cerebral, de amnesia y de un trastorno psicótico.

In the psychological context, the general concept of "simulation" covers *dissimulation* or *deception*, as well as *faking*, but the type of simulation we are concerned with here is normally defined as *malingering*, which refers to the conscious and deliberate invention of a physical or mental disorder, or the attribution of an existing incapacity to an accident or illness which was not actually its cause, in order to derive some personal benefit. The DSM-IV defines malingering as "the intentional feigning or exaggeration of physical or psychological symptoms, motivated by external incentives such as avoidance of work or military service, receipt of financial compensation, evasion of criminal prosecution, or procurement of prescription drugs. Under certain circumstances, malingering may constitute adaptive behaviour: for example, feigning illness while in captivity as a prisoner of war" (American Psychiatric Association, 1995, p. 698).

The concept is considered more closely by Resnick (1997), who distinguishes between *pure malingering*, simulation of a non-existent disorder; *partial malingering*, the conscious exaggeration of present symptoms or of a now-cured disorder; and *false imputation*, the erroneous attribution of real symptoms to a particular cause, due to non-deliberate self-deceit or a wrong interpretation of the situation.

Gorman (1982) also distinguished between the act and the state of malingering, insofar as the act implies wilfulness, an assertive attitude of desire and purpose, while the state, from a legal point of view, would be inherent to the person, due to their social condition or possible limitations.

On the other hand, the question arises as to whether malingering can, in itself, be the reflection of some mental disorder. This is fairly clearly the case in the so-called *factitious disorder* (American Psychiatric Association, 1994), in which the person intentionally feigns physical or psychological symptoms, with the aim of taking on the sick role, or in histrionic personality disorder, due to lack of control over one's manipulative behaviour; however, the conscious exaggeration of physical or mental complaints in order to achieve a financial, professional or similar objective may also be a neurotic behaviour, since nobody "in their right mind" would normally go to such extremes, nor choose such tortuous and painful routes, in the pursuit of potential benefits. Nevertheless, the law does not make such distinctions, and considers certain behaviours to involve clear intention on the part of the person carrying them out (Gorman, 1982).

The incidence of malingering after an accident is not well known, being estimated at between 1% and 50% of cases (Henderson, 1986; Miller & Catlidge, 1972), depending on whether the source of information is the claimant's lawyer or the insurance company. An influential factor here concerns the financial and employment conditions, since it has been observed that such malingering increases when redundancies are immi-

Correspondence: Mercedes Inda Caro, *Facultad de Psicología, Universidad de Oviedo, Plaza Feijóo, s/n, 33003 Oviedo, Spain. E-mail: indamaria@uniovi.es*

nent, and decreases when employees' financial or work situation improves. The disparity in estimations of the incidence of malingering may be due to the difficulty in distinguishing between those who totally invent the symptoms and those who exaggerate already-existing complaints –in whose case we could more properly speak of "patients".

Furthermore, the identification of malingerers and the study of their behaviour have traditionally been beset with difficulties. Thus, for example, in the context of compensation claims for the consequences of accidents, as in judicial processes in relation to diminished responsibility, it is found that people with access to information related to brain disorders and neuropsychological tests are more likely to able to influence their results (Coleman, Rapport, Millis, Ricker & Farchione, 1998; Youngjohn, Lees-Hayley & Binder, 1999). Research on malingering has shown training (such as the provision of information on the most common deficits associated with a brain lesion, or drawing attention to the presence of measures for detecting possible malingerers) to be another of the factors that can affect the validity of malingering indices. Studies carried out with university students, trained in simulating organic brain disorder, show that their performance is more similar to that of real patients than the performance of malingerers who are novices or who have no experience in the field of neuropsychological assessment. Nevertheless, the performance of such naïve malingerers is overwhelmingly poorer than that of persons with real brain lesions.

One of the many limitations observed in this type of research is the excessive use of samples of university students, who are asked to simulate brain damage (Strauss et al., 2002; Vickery, Berry, Inman, Harris & Orey, 2001), insofar as they are not comparable to true malingerers. Unlike simulators in experiments, real malingerers seek financial benefit from their "lesion", tend to have extensive knowledge of the problem through having undergone multiple assessments by a variety of experts (often at one- or two-week intervals), have observed patients with genuine disorders –thus learning passively and unconsciously–, and tend to have more time to prepare their assessments and examinations. Obviously, true malingerers do not acknowledge or admit their condition, which rules out their inclusion in studies. Despite such difficulties, in studies with analogue subjects it is possible to study simulation through the inconsistencies in repeated trials. In fact, it is quite difficult to maintain the same performance when large batteries of tests are employed, so that this method permits researchers to obtain a valid index for detecting possible malingering (Cullun, Heaton & Grant, 1991). Furthermore, on comparing patients involved in litigation over their lesions with patients with the same lesions but not involved in lawsuits, it is found that in the former group there is less consistency between their assessments than in the second group, the results of the later assessments being poorer than those of the initial assessments (Reitan & Wolfson, 1996, 1997).

DISORDERS THAT ARE FREQUENTLY THE OBJECT OF MALINGERING

Post-traumatic stress disorder (PTSD)

In order to establish PTSD it is necessary to carry out a meticulous description of the symptoms and the treatments previously applied and to carefully corroborate the veracity of the information. In the phase of obtaining information the clinician should be extremely careful not to provide any information to the person about the key symptoms of this disorder. Moreover, if the clinician begins the assessment by questioning the patient's responses, such an aggressive approach may affect the response style, and may lead the person to justify his or her injury by means of extreme symptoms. One of the disadvantages of clinical diagnosis is that it is based on patients' self-reports about subjective symptoms. Thus, the allegedly affected person's activity one week before the occurrence of the stressor is to be compared with their activity at the time of the assessment; on the basis of this, it is examined whether there is a reasonable relationship between the symptoms and the stressor, taking into account also the time elapsed since the stressor and onset of symptoms, and the relation between any previous disorder and the current symptoms. The psychologist should insist on being provided with a detailed description of the symptoms of the disorder. Malingerers may have extensive knowledge of the characteristic PTSD symptoms, but they normally fail to fit these symptoms to their everyday life, giving a description with little hard detail. Invented symptoms tend to be vague, and often quite contrived and unconvincing (Pitman, Sparr, Saunders & McFarlane, 1996). Another indicator of possible malingering is that the person minimizes other potential causes of their symptoms and exaggerates as the cause the accident or situation for which they are demanding compensation.

Resnick (1997) suggests that third persons should not be present during the assessment, for two reasons: first, because relatives or close friends present may be used as "independent" sources for corroborating the veracity of the symptoms; and second, because it is easier for the clinician to challenge a possible malingerer when the two are alone. Another aspect stressed by this author is the advantage of a friendly and cordial approach by clinicians on telling suspected malingerers that they think they might be exaggerating the symptoms, rather than being aggressive or trying to humiliate, by using witnesses, for example; with the latter approach, the person may refuse to admit it and become angry. Resnick sets some guidelines to be followed by the clinician who suspects simulated PTSD (Table 1), and a clinical decision model for determining the existence of malingering in the case of this disorder (Table 2).

Post-traumatic brain damage syndrome

This disorder is quite common today, largely as a consequence of occupational or road traffic accidents. It tends to be manifested through headaches, dizziness, anxiety, emotional instability, blurred vision, concentration deficit and memory problems. Of all the symptoms, the easiest to simulate are emotional ones. Post-traumatic brain damage syndrome can be confused with PTSD, since it is quite frequent after sustaining cranial-encephalic trauma. The two disorders have components in common, such as loss of memory of some element of the traumatic event, depressive symptoms (anhedonia, restricted affect, pessimistic attitude about the future), sleep alterations, irritability, concentration difficulties and intolerance of loud noises. However, some authors, such as Price (1994), maintain that it is not possible for the two disorders to coexist in the same person, since someone who has sustained a brain lesion with loss of consciousness will not be able to re-experience the traumatic event; hence, the mutually exclusive nature of the two disorders will justify the conclusion of malingering when the two are allegedly presented simultaneously.

Amnesia

The principal measures developed for detecting malingering in relation to memory disorders include very simple tests, which can be carried out correctly even by people with brain damage, and in which malingerers tend to show more deficits than true patients. A so-called *floor effect* comes into play when novice malingerers "overact" and commit many errors in these tests; however, Cercy, Schretlen and Brandt (1997) point to several problems with these techniques. First, people with experience in simulating amnesic symptoms are aware of this strategy and avoid performing too badly in the tests. Secondly, despite the apparent simplicity of the tests, some patients with real brain damage or with neuropsychiatric disorders have considerable difficulties in carrying them out correctly.

A new current is developing for the detection of malingering in relation to amnesic disorders, represented by analysis of the *proactive interference phenomenon*. This phenomenon occurs when previously learned information interferes with the acquisition or subsequent recall of a new message, and it is analyzed, for example, by means of the paradigm of learning a word list. Proactive interference is reflected in a decline in memory for the words in it as the list progresses, so that the first words learned

TABLE 1
THRESHOLD MODEL FOR THE EVALUATION OF MALINGERING IN POSTTRAUMATIC DISORDERS (RESNICK, 1997)

Any combination of the following criteria:
1. Poor work record
2. Prior 'incapacitating' injuries
3. Markedly discrepant capacity for work and recreation
4. Unvarying, repetitive civilian dreams
5. Antisocial personality traits
6. Overly idealized functioning before the trauma
7. Evasiveness
8. Inconsistency in symptom presentation

TABLE 2
CLINICAL DECISION MODEL FOR ESTABLISHING MALINGERED PTSD (RESNICK, 1997)

A. Understandable motive to malinger PTSD
B. At least two of the following criteria:
1. Irregular employment or job dissatisfaction
2. Prior claims for injuries
3. Capacity for recreation, but not for work
4. No nightmares or, if nightmares, exact repetitions of the civilian trauma
5. Antisocial personality traits
6. Evasiveness or contradictions
7. Noncooperation in the evaluation

C. Confirmation of malingering by one of the following criteria:
1. Admission of malingering
2. Unambiguous psychometric evidence of malingering or strong corroborative evidence of malingering

interfere in the coding and storage of subsequent ones. Interference is greater when the new information is very similar to that already stored (as in the case of words from the same semantic category) than when a different category is introduced; in the latter case there is some recovery of memory processes (Wickens, 1970). The effect of proactive interference has been shown in patients with real brain lesions, but not in people attempting to simulate memory impairment, who recall the later words of the list better than the earlier ones. This phenomenon, however, has not been confirmed by other authors, who have failed to find differences of degree between malingerers and patients with genuine cerebral damage (Baker, Hanley, Jackson, Kimmance & Slade, 1993). The relevance of proactive interference as a detector of malingering is based on the assumption that it is an automatic cognitive process, outside the conscious control of the subject.

Baker and cols. (1993) also explored the potential influence of a distractor on recall of a set of stimuli when it appeared between their presentation and the recall task; however, they found no significant differences when recall of the items was required after a 20-second interval in which subjects had to count backwards. False patients, on the other hand, performed far worse in this test.

Other studies have focused on the detection of malingering through tests of implicit memory (García Domingo, Gregredo López & Fernández Guinea, 2004). The performance of amnesic persons is generally closer to that of normal persons in tests with the *priming* effect, and in tasks that do not require explicit recall of the learned episode. For example, when patients with amnesia process a series of words without being told that they will later be asked to remember them, and are subsequently shown the root of the word or fragments of it, the probability of their recalling the word correctly is very high. This *priming* phenomenon is considered to be controlled by independent neurocognitive processes. Wiggins and Brandt (1988) suspected that malingerers would perform these implicit memory tests more poorly than true patients and, indeed, they found malingerers' performance to be relatively poorer, though they did not find statistically significant differences.

Another procedure for detecting malingerers has involved assessment of the so-called *feeling-of-knowing*, or a person's sensation of having partial recall: in other words, whether the person is aware of suffering from amnesia. However, some authors consider this indicator to be limited, given the variability among true amnesic patients. Even so, it has been found that people who simulated memory disorders and who had scored lower in a forced-choice test showed low levels in *feeling-of-knowing* (Schacter, Harbluck & McLachlan, 1984).

Psychosis

The prevalence of simulated psychosis is unknown, though Resnick (1984) considers that, given the trend towards deinstitutionalization, it could be on the increase, since thousands of chronic patients, who would prefer to live in a more protected environment, currently find themselves socially marginalized. With the drastic cutbacks in social programmes and improvements in hospital conditions, people with mental disorders may exaggerate their symptoms in order to obtain medical help; such behaviour would be comparable to that of patients with schizophrenia, who display a remarkable ability to appear healthy or sick depending on their current objectives (Rogers, Kropp, Bagby & Dickens, 1992). Simulation of a psychotic disorder may occur for a variety of reasons: avoidance of responsibility by persons involved in judicial proceedings; avoidance of military service or of postings to dangerous places (no longer applicable in Spain); obtaining financial benefit due to psychological injury or effects; release from standard prison conditions (simulation of a psychotic state to obtain transfer to a hospital, in order to gain easier access to drugs or improve the chances of escape).

Specialists lament the lack of diagnostic criteria for establishing the existence of malingering in these cases. Resnick (1997), however, suggests some principles to be taken into account by the clinician who suspects a case of malingering. Thus, with regard to auditory hallucinations, suspected malingerers should be asked about the strategies they employ for reducing the voices or making them disappear. In addition to the fact that genuine patients tend to present a reduction in this type of hallucination when the schizophrenia is in remission, while in acute outbreaks they occur with great frequency, the coping strategies used by patients with schizophrenia include specific activities, such as working or watching television, changing position (e.g., lying down or walking around), talking to a friend or relative, or rapidly taking one's medicine; in general, they find that their hallucinations tend to decrease when they are involved in some activity. These spontaneous actions and the corresponding mitigating effect on the hallucinations should

be assessed in suspected malingerers, since, if they do not have profound knowledge of the illness, they will fail to provide such information during the interview. Genuine hallucinations are characterized by a wide range of nonsensical murmurs and cries; on the other hand, the rhythm of discourse is normal. In contrast, malingerers sometimes refer to the content of their hallucinations in a contrived and over-complicated way. With regard to visual hallucinations, Resnick suggests that malingering should be suspected when their content is dramatic or atypical.

As far as delusions are concerned, malingerers report their sudden appearance, when it is well known that "real" delusions are built up over months or years, until they become systematized. When delusional ideas appear, they tend to have little influence on the patient's everyday life, even though the patient is convinced of their veracity. In the assessment of whether or not a delusion is genuine, Resnick points out the importance of considering its content. Feigned delusions tend to be persecutory, or of grandeur, but are rarely self-deprecatory. Moreover, the behaviour of malingerers is not usually in accordance with the content of the supposed delusion, whilst in persons with genuine psychosis, its behavioural relevance is greater.

Another symptom that persons with psychotic disorder often present is mutism. In malingerers, mutism may appear as an isolated symptom or as part of a more general simulated psychosis. Catatonic behaviour or waxy flexibility are very difficult to maintain for prolonged periods, so that a way of determining whether a person is feigning is to see how they react when pricked on the back with a pin. Those with genuine catatonia will respond in the same way regardless of whether they see the painful stimulus in advance; malingerers, on the other hand, will respond differently depending on whether they anticipate the stimulus or not: if they see the examiner approaching them with the pin, they will present a small reaction, having previously tensed their muscles; if they are pricked without having seen the pin they will present less muscular contraction and pupillary dilation.

In the case of conversion disorders, it is more difficult to detect malingering. Resnick argues that the distinguishing criterion is whether the mutism behaviour is under the person's voluntary control. Knowledge of the precise details of how the person came to stop speaking is very important, according to Resnick. People with conversion disorders are usually capable of writing and whispering, and tend to have a history of other conversion symptoms, as a result of a dissociative disorder, for example; on the other hand, in malingerers it is common to find a history of antisocial behaviour with lying, and a criminal record.

As regards simulation of psychotic depression, it is well known that diurnal variation forms part of its clinical expression, so that the genuine patient presents greater severity of symptoms and more dysphoric mood states in the morning and some improvement towards the end of the day. Such clinical fluctuation is less likely to be reported by malingerers, on lacking profound knowledge of the disorder.

ASSESSMENT METHODS

Procedures for the assessment of malingering, in the clinical context, have been based on the use of conventional neuropsychological measures and of specifically designed tests.

Within the first option, researchers have analyzed: performance curves in tasks of varying difficulty (Baker et al., 1993; Frederick, Crosby & Wynkoop, 2000; Tehula & Sweet, 1996); correct responses in recall and recognition tasks and tasks that require subjects to discriminate between two types of stimulus (Coleman et al., 1998; Slick, Iverson & Green, 2000; Suhr & Gunstad, 2000; Sweet et al., 2000); memory tasks (Davis, King, Bajszar & Squire, 1995; Hanley, Baker & Ledson, 1999); digit span (Strauss et al., 1999; Suhr, Tranel, Wefel & Barrash, 1997); comparison of attention and memory indices (Mittenberg, Azrin, Millsaps & Heilbronner, 1993); and semantic knowledge (Mittenberg, Theroux-Fichera, Heilbronner & Zielinski, 1995). Although instruments of this type are considered optimum for detecting possible cases of malingering, it is also deemed necessary to apply complementary tests to improve the validity and reliability of the results. Lezak (1995) lists the following classic neuropsychological tests for the detection of malingerers:

- The Bender Test, with the recommendation of carrying out a retest several days after the first assessment (since the subject will forget what the response patterns were), and inverting the order of the cards.
- The Benton Visual Retention Test, in which malingerers make more distortion errors than patients with brain lesions, but not more omission errors.
- The Halstead-Reitan Battery (including the WAIS), in which malingerers perform worse on the tests than patients with lesions, except in the cases of the Category Test, the Tactile Performance Test and part B of the Trail Making Test.

- The Minnesota Multiphasic Personality Inventory (MMPI), in which malingerers also obtain poorer profiles than genuine patients.
- The Porch Index of Communicative Ability (PICA), for malingering of aphasic disorder.

Among the tests designed specifically for assessing malingering, two methodological lines can be identified. One is based on the so-called *symptom validation paradigm* (Pankratz, Fausti & Peed, 1975), originally designed for assessing deficits in sensory functioning, and later extended to the detection of simulation of memory-related lesions (Binder & Willis, 1991; Frederick & Foster, 1991; Iverson, Franzen & McCracken, 1991; Pankratz, 1983). This paradigm involves the administration of forced-choice tests with two response options, whose results are based on probabilities (Slick, Hopp, Strauss & Thompson, 1997; Tombaugh, 1996), and which set confidence intervals, above or below which scores are considered as indicative of simulation or exaggeration of symptoms, and cut-off points for the selection of responses. For example, persons who are not trying to simulate should obtain at least 50% of correct responses, which is the result that would be expected from someone responding at random. This was the cut-off criterion initially employed, but it was found in trials that normal persons pretending to be malingerers did not score below the response levels expected by chance, though they did make more errors than genuinely sick and honest patients. Therefore, it was decided to establish cut-off points in relation to the performance expected of a person with a real lesion and no intention to exaggerate or simulate.

Some researchers have begun to explore the utility of concealed measures, obtained from "objective" responses made by suspected malingerers, which cannot be manipulated by these persons or "improved" with successive assessments. An example of this is the computerized version of the *Portland Digit Recognition Test*, by Rose, Hall and Szalda-Petree (1995), which includes a measure of subject's response latency. These authors found that the incorporation of this measure into the original version created by Binder and Willis (1991) improved the test's sensitivity in the identification of possible malingerers.

The second methodological line in tests designed specifically for the assessment of malingering involves the study of the type of response the patient makes; for example, the way in which the patient reads very simple words or counts the number of dots appearing on a screen (Boone et al., 2000; Strauss et al., 2002). An example of this would be the *Dot Counting Test* (Binks, Gouvier & Waters, 1997), in which subjects are presented with a series of cards with grouped and ungrouped dots and asked to count the number of dots they see on the screen, scores being based on number of hits and time employed in counting the stimuli.

ASSESSMENT BY MEANS OF INTERVIEWS AND SELF-REPORTS

Another form of detecting malingering consists in assessing the behavioural symptoms of the problem. Initial approximations were made by means of tests whose specific objective was not the detection of deceit, but which included some subscales for measuring the validity of the instrument. The first of these, and the most well-known, is the *Minnesota Multiphasic Personality Inventory* (MMPI), whose F Scale has the purpose of detecting atypical response styles. With this scale, however, various problems were detected, such as overlap between scores obtained by genuine patients and those obtained by possible malingerers, and the poor sensitivity of this scale for detecting specific malingering situations, such as the simulation of memory disorders. The DM subscale of Cattell's 16 PF has also been the object of criticism in this context.

The *M Test* (Beaber, Marson, Michelli & Millis, 1985) was the first instrument developed with the primary objective of detecting possible malingerers. However, validation studies have also called this test into question, with Hankins, Barnard and Robbins (1993) arguing that it seems rather to detect persons that present cognitive deficit or deterioration. In an attempt to improve the test, Rogers and cols. developed a new scoring system, obtaining optimum results in the differentiation between psychiatric patients from a prison and a hospital (Rogers, Bagby & Gillis, 1992). Smith, Forum and Schinka (1993), on the other hand, failed to confirm these results with a similar population.

The *Malingering Scale* (Schretlen, 1986) constitutes another attempt to construct a test for detecting malingerers by means of psychometric assessment. This instrument consists of two scales: the *malingering retardation*, or *MgR* scale, and the *malingering insanity*, or *MgI* scale. However, Smith and Burger (1997) point out that studies developed for validating the test have methodological deficits, related to sample bias, and that the test itself has practical drawbacks, such as its length (150 items) and

the need for it to be applied by an experienced assessor.

Rogers and cols. developed the SIRS (*Structured Interview of Reported Symptoms*) with the aim of unmasking those who are feigning or exaggerating a mental disorder (Rogers, Gillis & Bagby, 1990; Rogers, Gillis, Dickens & Bagby, 1991; Rogers, Kropp et al., 1992). The SIRS is an interview with 172 questions, distributed in 8 primary scales and 5 additional or complementary scales. The former are used for distinguishing between honest respondents and malingerers, and for assessing response styles. These primary scales explore: Rare Symptoms (RS) (genuine, but uncommon); Improbable and Absurd Symptoms (IA); Symptom Combinations (SC) (referring to the low probability of two genuine symptoms being presented simultaneously); Blatant Symptoms (BL); Subtle Symptoms (SU) (referring to symptoms seen by patients as problematic, when in fact they are not); Selective Symptoms (SEL); Symptom Severity (SEV); and Reported versus Observed Symptoms (RO).

With the complementary scales the clinician can make an interpretation of the subject's response styles. This second block consists of: Direct Appraisal of Honesty (DA); Defensive Symptoms (DS); Symptom Onset (SO); Overly Specified Symptoms (OS); and Inconsistency of Symptoms (INC).

The questions can be grouped in three categories: (a) detailed questions, aimed at exploring the severity of specific symptoms; (b) repeated questions, for control purposes in relation to responses to the questions from section a; and (c) general questions, aimed at examining patterns of symptoms and psychological problems.

In developing the SIRS, Rogers (1984) reviewed the literature to identify strategies of possible utility for the detection of malingerers, selecting those that met the following criteria: (a) relevance for detecting malingering of a mental illness, as opposed to other forms of deceit, and (b) ease of standardization. On the basis of these criteria, Rogers selected five strategies, from which he generated 330 questions to make up the first version of the SIRS. The scales were formally constructed following two steps: on the one hand, based on the agreement among eight experts, the apparent and descriptive validity of the proposed scales were sought, the items being assigned to the strategy they believed most appropriate, so that when at least five of the eight experts coincided with Rogers' classification the item was placed in the corresponding scale; and on the other hand, the item-scale correlations were calculated, those items that failed to correlate with the assigned scale being eliminated. The alpha coefficients of the scales were between 0.66 and 0.92, with a mean of 0.86 (Rogers, 1997).

The results for each one of the scales are classified in four categories: honest, indeterminate, probable malingering and definite malingering. The person is considered to be attempting to deceive if the score on three or more of the primary scales is in the range of *probable malingering*; or if the total SIRS score (the sum of those for the general questions and the detailed questions) exceeds 76. The person is considered to acting honestly if the score on six or more of the primary scales is in the range of *honest*, or if the global score is 71 or less.

Finally, the SIMS (*Structured Inventory of Malingered Symptomatology*) (Smith & Burger, 1997) is another instrument for the assessment of malingering, involving the self-report of 75 dichotomous items (true/false), grouped in five scales developed for detecting possible deception in the five most common clinical conditions of malingering: low intelligence, affective disorders, neurological damage, psychosis and amnesia. A total score is obtained from the five scales. Items were obtained from two different sources: first, already-existing instruments, such as the MMPI, SIRS and WAIS-R, which have shown some utility in the detection of possible malingering (these items were modified in order to increase their sensitivity in the detection of specific malingering situations); and second, the qualitative characteristics of malingerers (Resnick, 1984; Rogers, 1984; Seamons, Howell, Carlisle & Roe, 1981).

CONCLUSIONS

Malingering, deception or feigning potentially occur in all types of somatic illnesses and mental disorders. It is therefore necessary to use different assessment procedures for unmasking persons presenting or exaggerating a wide range of symptoms; it is by no means the same to simulate physical damage, such as a brain lesion, as it is to feign psychological damage, such as a mental disorder. Whatever the nature of the symptoms, it is often advantageous for a psychologist to participate in the assessment of their authenticity –particularly when the person's alleged problems affect the cognitive functions, such as attention or memory, and are accessible to neuropsychological assessment. Despite the fact that medical examinations can rule out organic brain damage, patients may often report problems in their everyday life when it comes to driving, remembering things, and so

on, and these are the factors mentioned by expert witnesses in court cases, and on which it is necessary to make a decision.

Credibility in relation to the genuine or simulated nature of symptoms or of these testimonies, insofar as it involves subjective opinions, is not particularly accessible to scientific study. Nevertheless, what can be studied is the validity of the symptoms or the clinical condition the person presents, in order to determine with scientific criteria a probable situation of malingering or exaggeration. Thus, it is suggested to undertake a multi-factor approach for determining the existence of a malingering situation; for this, it is necessary to: (1) determine the severity of the damage, through the verification of different symptoms; (2) assess the patient by means of standardized tests; (3) consider alternative psychological or medical diagnoses for explaining the cause of the symptoms adduced by the person; (4) use tests suitable for the demographic characteristics of the subjects assessed; and (5) use, simultaneously, neuropsychological tests and specific validity indicators for determining possible feigning of symptoms.

Esbec Rodriguez and Gómez-Jarabo (1999) have described, for example, up to twenty characteristics that can indicate malingering of a mental disorder, two of which stand out as the most important: the presence of some clear external benefit or gain due to the presentation of these symptoms, and verification that the subject had previously presented similar symptoms to those allegedly suffered at present.

Forensic Psychology, therefore, both in our own country and elsewhere, is faced with an important challenge: to determine, on a scientific basis, the validity of the testimonies and alleged symptoms of persons involved in judicial proceedings or who have been the victims of accidents or violence. In particular, it is necessary to develop structured and standardized procedures that permit well-founded judgements on the possibility of malingering. In the pursuit of this aim, the procedures developed by Arce and Fariña (2005) constitute a sound example to be followed.

REFERENCES

American Psychiatric Association. (1995). *Manual Diagnóstico y Estadístico de los Trastornos Mentales (4th ed.)*. Barcelona: Masson.

Arce, R. & Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG). *Papeles del Psicólogo, 26*, 59-77.

Baker, G. A., Hanley, J. R., Jackson, H. F., Kimmance, S. & Slade, P. (1993). Detecting faking of amnesia: Performance differences between simulators and patients with memory impairment. *Journal of Clinical and Experimental Neuropsychology, 15*, 668-684.

Beaber, R. J., Marson, A., Michelli, J. & Millis, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. *American Journal of Psychiatry, 142*, 1478-1481.

Binder, L. M. & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*, 175-181.

Binks, P. G., Gouvier, W. D. & Waters, W. F. (1997). Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology, 12*, 41-46.

Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., et al. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b test. *Archives of Clinical Neuropsychology, 15*, 227-242.

Cercy, S. P., Schretlen, D. & Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomena. In R. J. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 85-107). New York: Guilford Press.

Coleman, R. D., Rapport, L. J., Millis, S. J., Ricker, J. H. & Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 20*, 201-210.

Cullun, C., Heaton, R. & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H. O. Doerr & A. S. Carlingf (Eds.), *Forensic neuropsychology: Legal and scientific bases* (pp. 141-171). New York: Gilford Press.

Davis, H. P., King, J. K., Bajszar, J. H. & Squire, L. R. (1995). *Colorado Malingering Test Package, version 2.0*. Colorado Springs, CO: Colorado Neuropsychology Tests.

Esbec Rodríguez, E. & Gómez Jarabo, G. (1999). Signos de sospecha en la simulación de los trastornos mentales. *Revista Española de Psiquiatría Forense, Psicología Forense y Criminología, 8*, 35-44.

Frederick, R. L., Crosby, R. D. & Wynkoop, T. E. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology, 15*, 281-300.

Frederick, R. L. & Foster, H. G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment, 3*, 596-602.

García Domingo, L., Gregredo López, L. & Fernández Guinea, S. (2004). Evaluación de la simulación de problemas de memoria dentro del ámbito legal y forense. *Revista de Neurología, 38*, 766-774.

Gorman, W. F. (1982). Defining malingering. *Journal of Forensic Sciences, 27*, 401-407.

Hankins, G. C., Barnard, G. W. & Robbins, M. L. (1993). The validity of the M Test in a residential forensic facility. *The Bulletin of the American Academy of Psychiatry and Law, 21*, 111-121.

Hanley, J. R., Baker, G. A. & Ledson, S. (1999). Detecting the faking of amnesia: A comparison of the effectiveness of three different techniques for distinguishing simulators from patients with amnesia. *Journal of Clinical and Experimental Neuropsychology, 21*, 59-69.

Henderson, J. (1986). Psychic trauma claims in civil and administrative law. *Paper presented at the 17th American Academy of Psychiatry and the Law Meeting*, Philadelphia, PA.

Iverson, G. L., Franzen, M. D. & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior, 16*, 667-676.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Miller, H. & Catlidge, N. (1972). Simulation and malingering after injuries to the brain and spinal cord. *Lancet, 1*, 580-585.

Mittenberg, W., Azrin, R., Millsaps, C. & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. *Psychological Assessment, 5*, 34-50.

Mittenberg, W., Theroux-Fichera, S., Heilbronner, R. & Zielinski, R. E. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology, 11*, 271-281.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57*, 367-372.

Pankratz, L., Fausti, A. & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43*, 412-422.

Pitman, R. K., Sparr, L. F., Saunders, L. S. & McFarlane, A. (1996). Legal issues in posttraumatic stress disorder. In B. A. van der Kolk, A. C. McFarlane & L. Weisaeth (Eds.), *Traumatic Stress* (pp. 378-397). New York: Guilford Press.

Price, K. P. (1994). Posttraumatic stress disorder and concussion: Are they incompatible? *Defense Law Journal, 43*, 113-120.

Reitan, R. M. & Wolfson, D. (1996). The question of validity of neuropsychology test scores among head-injured litigants: Development of a Dissimulation Index. *Archives of General Neuropsychology, 2*, 573-580.

Reitan, R. M. & Wolfson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: Development of the Retest Consistency Index. *Clinical Psychologist, 11*, 69-76.

Resnick, P. J. (1984). The detection of malingered mental illness. *Behavioral Sciences and the Law, 2*, 20-38.

Resnick, P. J. (1997). Malingering of posttraumatic disorders. In R. J. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 130-152). New York: Guilford Press.

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law, 2*, 93-112.

Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception*. New York: Guilford Press.

Rogers, R., Bagby, R. M. & Gillis, J. R. (1992). Improvements in the M Test as a screening measure for malingering. *The Bulletin of the American Academy of Psychiatry and Law, 20*, 101-104.

Rogers, R., Gillis, J. R. & Bagby, R. M. (1990). Cross validation of the SIRS with a correctional sample. *Behavioral Sciences and the Law, 8*, 85-92.

Rogers, R., Gillis, J. R., Dickens, S. E. & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the SIRS. *Psychological Assessment, 3*, 89-96.

Rogers, R., Kropp, P. R., Bagby, R. M. & Dickens, S. E. (1992). Faking specific disorders: A study of the Structured Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology, 48*, 643-648.

Rose, F. E., Hall, S. & Szalda-Petree, A. D. (1995). Computerized Portland Digit Recognition Test: The measurement of response latency improves the detection of malingering. *Clinical Neuropsychologist, 9*, 124-134.

Schacter, D. L., Harbluck, J. L. & McLachlan, D. R. (1984). Retrieval without recognition: An experimental

analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior, 23*, 593-611.

Schretlen, D. (1986). *Malingering: Use of a psychological test battery to detect two kinds of simulation*. Ann Arbor, MI: University Microfilms International.

Seamons, D. T., Howell, R. J., Carlisle, A. L. & Roe, A. V. (1981). Rorschach simulation of mental illness and normality by psychotic and nonpsychotic legal offenders. *Journal of Personality Assessment, 45*, 130-135.

Slick, D. J., Hopp, G., Strauss, E. & Thompsom, G. (1997). *Victoria Symptoms Validity Test*. Odessa, FL: Psychological Assessment Resources.

Slick, D. J., Iverson, G. L. & Green, P. (2000). California Verbal Learning Test indicators of suboptimal performance in a sample of head-injury litigants. *Journal of Clinical and Experimental Neuropsychology, 22*, 509-579.

Smith, G. P., Borum, R. & Schinka, J. A. (1993). Rule-out and rule-in scales for the M Test for malingering: A cross-validation. *The Bulletin of the American Academy of Psychiatry and Law, 21*, 107-110.

Smith, G. P. & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *The Bulletin of the American Academy of Psychiatry and the Law, 25*, 183-189.

Strauss, E., Hultsch, D. F., Hunter, M., Slick, D. J., Patry, B. & Levy-Bencheton, J. (1999). Using intraindividual variability to detect malingering in cognitive performance. *Clinical Neuropsychologist, 14*, 420-432.

Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. & Hultsch, D. F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology, 17*, 423-444.

Suhr, J. A. & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15*, 415-424.

Suhr, J. A., Tranel, D., Wefel, J. & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology, 19*, 500-514.

Sweet, J. J., Wolfe, P., Sattleberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., et al. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology, 15*, 105-114.

Tehula, W. N. & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *Clinical Neuropsychologist, 10*, 104-116.

Tombaugh, T. N. (1996). *Test of memory malingering*. New York: MHS.

Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J. & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16*, 45-73.

Wiggins, E. C. & Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behavior, 12*, 57-78.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*, 1-15.

Youngjohn, J. R., Lees-Hayley, P. R. & Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingering. Archives of Clinical Neuropsychology, 14, 511-515.

Papeles del Psicólogo, 2005. Vol. 26, pp. 109-114

DECEPTION AND LYING IN PSYCHOLOGICAL DISORDERS AND THEIR TREATMENTS

Manuel Porcel Medina* and Rubén González Fernández**
**Department of Employment Advice, General Workers' Union*
***Private Centre for Psychological Attention*

This article attempts to examine the phenomena of lying, deception and self-deception as directly related to psychological problems and their treatments. We shall see how, in many cases, deception and self-deception are no more than a psychological or fictitious covering for life problems. But given that psychological problems can only be neutralized through psychotherapy, we shall see that all self-respecting psychotherapies should incorporate lying in their therapeutic games. And the therapist will put this psychotherapeutic lying into practice in a highly theatrical way, by means of the technique known as "paradoxical intention". Indeed, it is this technique, bound up with deception, that can best counteract self-deception in a psychologically disturbed client. Paradoxical intention will work when adjusted to the client's movements, shaping those movements on the "road" the client is moving along at each moment.

Este artículo pretende examinar los fenómenos de la mentira, el engaño y el autoengaño como términos que están directamente relacionados con los problemas psicológicos y sus tratamientos. Veremos que, en numerosas ocasiones, el engaño y el autoengaño no son más que la cobertura psicológica o ficticia de los problemas de la vida. Pero como los problemas psicológicos sólo pueden neutralizarse a través de psicoterapia, veremos que toda psicoterapia que se precie deberá incorporar en sus juegos terapéuticos a la mentira. Y el terapeuta ejercitará esta mentira de la psicoterapia, de una manera muy efectiva, a través de la técnica que conocemos como "intención paradójica". En efecto, será esta técnica intrincada en la mentira la que pueda contrarrestar la propia mentira del cliente aquejado de un problema psicológico. La intención paradójica funcionará cuando se administre acompasada con el movimiento del cliente, moldeando dicho movimiento en el camino que el propio cliente estará transitando en cada momento.

There are certain pithy phrases which attempt to give a concise definition of man: "man is an animal that uses tools"; "man is an animal with the capacity for language"; "man is an animal that eats bread", and so on. Our contribution to this catalogue of phrases is: "man is an animal that lies". Clearly, other animals use deceit for survival, but humans are perhaps the only creatures that use lies in a reflective way, that is, the only ones capable of using lies in relation to their person, to their identity. If the chameleon camouflages its body, humans would be capable of camouflaging the very depths of their guts. Suggestion is a prodigy particular to the human mind. Only man can see ghosts.

At the risk of exceeding the parsimonious objectives of this article, we should like to stress the enormous importance of lying, insofar as it can be associated with suggestion, since hidden behind is the foundation of psychology itself: the self, reflectiveness.

Correspondence: Rubén González Fernández.
E-mail: rgonzalez@correo.cop.es

THE TRUTH OF LIFE AND THE VITAL TEXTURE OF PSYCHOLOGICAL PROBLEMS

We shall discuss this aspect with the help of a metaphor, and for two basic reasons. First, to aid understanding and to support our explanation. And second, in honour of the psychotherapy we ourselves practice –an updated and adjusted version of what is generally known as Acceptance and Commitment Therapy (ACT)–, since one of the basic principles of this therapy is the conviction that metaphors are genuine therapeutic tools in the clinical context. They are used to help the client see a series of concepts and phenomena that would otherwise be difficult to discern.

The metaphor we shall use is that of "the road". And we do not mean a poetic road, a quiet country road with its charm and mystery. That's not the kind of road we are thinking of. The road we shall use as a metaphor is a modern, urban road, with its roundabouts, its shopping centres on either side, its busy crossings with traffic lights, zebra crossings, give way signs, and so on.

And it has to be a modern, urban road because psychological problems have emerged in a modern and urban context that is none other than the context of the city. In the societies of old there was little room for psychological problems, since life was highly normalized. Society was closed, and that meant that the way of life and the problems arising from it were contained by strict boundaries. The city, on the other hand, has its very origins in a crossing of ways of life. We could say that it is no longer the family that regulates ways of life, but rather "the market". Everything is commercialized now: food, clothes, everyday goods and products. But there is also a trade in ways of life, and we are offered different professional, family and leisure alternatives, which are not only alternative but also, in many cases, incompatible and contradictory. And the problem with having access to different ways of life is that the individual begins to exercise the "responsibility of choice". We might say –looking ahead to what will be our principal thesis– that psychological problems are related, more or less directly, to this responsibility of choice.

Returning to our metaphor of the modern, urban road, let us imagine that *life* is a roundabout. A roundabout with the function of distributing the traffic. It gives order to a crossing of ways and it is the part of the urban road that helps us to change direction.

What would happen if there were no roundabouts? We would probably leave one road and take another without further delay, directly, without a waystation. Haven't you ever gone round our roundabout more than once to sort out your ideas about which direction to take? It's a kind of extra-decisional time. But what if we stay on the roundabout indefinitely without taking any exit? These are for us *psychological problems.* It is a "life jam" in the decision about which way to take at a particular moment. We have several alternatives, all with their advantages and disadvantages. Some offer a very easy passage but lead to unattractive places. Others are very hard, with a lot of traffic, with traffic lights, but lead us to much more inviting places. But we make no decision, we are "jammed". And while just one of us is jammed on the roundabout, there's no great problem. The problem grows when the roundabout gets jammed up and no longer allows others to use it properly. This is when the individual psychological problem turns into a problem of social dimensions. And the problem also increases when, being permanently on the roundabout, we get nowhere –we don't do our job in the city properly, we fail to meet our family obligations, and so on.

For this situation an entire circulatory network, alternative and outside of life, has been created, which permits us to continue driving, but with no destination. We drive in circles because we have to keep moving. Our sole objective is to remain in "pause" for the time we need to get back onto our ordinary urban road.

This alternative circulatory network, sterile and removed from life, is *psychologization*. And we are not referring only to the psychologization practised through the word, but also that which makes use of drugs, of flowers –read "Bach Flower Essences" for example–, of magic, of futurology, and so on. This whole network is at the service of those who got stuck on the roundabout indefinitely, and were unable to make a life decision.

Psychotherapy has become an intermediate institution (Pérez Álvarez, 1999) between institutions that fail. When people break down between various life alternatives, none of which satisfy, there emerges a concealed or covered way that makes sense at the time but ultimately does not, which is "psychological problems". The psychotherapeutic institution provides the necessary coverage for psychological problems by giving a functional explanation. A person can live between the conflicting alternatives of wanting to be slim and bingeing on food. The breakdown situation for this person would be that in which they put off indefinitely the decision between one way and the other: gorging themselves on food without worrying about the consequences, or trying to eat appropriately. The intermediate way (the roundabout of psychopathological life) tends to be to gorge oneself and vomit. The certainty that taking one road or the other depends on oneself is clearly unbearable. And this is the attraction of the third way, the dead and empty road of labelling this absence of personal decision a "psychological problem". This is the origin of a whole normative framework that protects and explains such irrational behaviour as gorging oneself and vomiting it all out. The psychological problem is called bulimia, and the professional who has to get rid of it is the psychologist or psychiatrist. From that point on, the only person who could turn the situation around –the person with bulimia– becomes subjugated to the psychotherapeutic institution, which now does everything for them. It is precisely this loss of responsibility that turns life problems, temporary log-jams, into psychological problems, into dead-end streets, without structural exits.

But of course, paradoxical as it may seem, it is only through psychotherapy that the lost sanity can be re-established. It is by means of psychotherapy that we are to turn psychological problems back into life problems, into challenges or difficulties that require continual decision-making by the person experiencing the problem. When decision-making becomes encysted –and we get stuck on the roundabout– we are feeding a problem until it becomes psychological. Put succinctly, we must depsychologize the client from psychology itself. Depsychologizing means, in this context, removing the psychological covering –external to the client and unable to be confronted– from life problems. Decisions must be made by clients, and it is precisely the psychologist's duty to avoid being tempted to make those decisions for them. Using the framework that permits a full and comprehensive explanation of psychological problems, "author-actor" (Quiroga Romero, 1999), we can state that psychotherapy, from the perspectives we are concerned with here (contextual, behavioural, ontological) is simply the attempt to move clients along the gradient from the irresponsibility and indecision of the life they are living, as mere actors playing a role, to responsibility and real contact with that life, becoming the actual author of it. How to achieve this change is what we shall try to explain below.

SELF-DECEPTION AS MODULATOR OF PSYCHOLOGICAL DISORDERS

A relevant aspect in psychopathological casuistry is revealed through the phenomenon of self-deception. If deception in general is a near-universal element of social interaction in modern societies, from the political and economic spheres to the most intimate contexts, self-deception might be seen as the de-generated extension that ends up insidiously impregnating personal consciousness and will.

Self-deception could be defined as way of leading our life when not only are we ignorant of what the chosen direction involves, but when, above all, we are ignorant or try to be ignorant of the fact that we have irremediably taken a route that brings with it a series of consequences. This is the meaning Plutarch gave to the term, when he said that self-deception was something more than the inability to recognize that we know nothing of many things, since in the end, the most dramatic thing is that we do not know what we are. If uncertainty and insecurity paralyze us, and out of prudence we decide to stagnate, we might ask ourselves whether paralysis might not also be an option involving risks, and therefore sometimes an imprudent option.

A lie can have different variants. It can be innocent or humorous, it can be somewhat perverse, and even kind or useful. Self-deception, on the other hand, without prejudice to its consideration as innocent, humorous, perverse, compassionate and useful, is not of one type alone, but rather a little of all of them. Deception involves a conscious objective, but self-deception is unconscious –we do not know what we are doing; as Oscar Wilde so pertinently remarked, "she is a veil, rather than a mirror" (Wilde 1889). Psychological disorders display this peculiarity in the majority of cases. Thus, a person affected by anorexia is often ignorant of their fear of public rejection, focusing their efforts on slimming or a struggle with their body. Neurotics with compulsive behaviours are unaware that, concealed behind their need to wash their hands constantly or their dread of contamination, is their stubborn refusal to accept the necessarily uncertain nature of life.

The main challenge for the psychologist tends to be to clearly reveal the real problem, which is generally hidden from the person suffering from it.

In general, people with psychological problems suffer because of something that they themselves exclude or push away, but to which they are nevertheless committed. Conscious will, we might say, is given over to the attitude of struggling reflectively with the problematic psychological elements, and this distraction keeps the person from acknowledging and perhaps being able to overcome the real problem affecting their life. As the philosophy of Acceptance and Commitment Therapy, and other, previous philosophies and authors have emphasized, what underlies a person's striving to control emotional and cognitive symptoms (which they paradoxically feed) is existential or vital (experiential) avoidance, a difficulty to accept things that cannot be changed (Luciano & Hayes, 2001).

Here, the phenomenon of self-deception emerges as something crucial, in the sense that the effort of concentrating on the psychological elements ends up concealing the substantial elements of an unresolved personal conflict (Fuentes Ortega, 1994), and this in turn confers a psychic character on a problem that only personal confrontation can finally resolve. A person complaining of

depression can thus hide their responsibility to confront the pain, suffering and sorrow behind their insistence on staying in bed, on remaining apathetic and scarcely active. But it is only when they abandon themselves to continual self-inflicted torture, reproaching themselves for their state of depression, that the self-deception or a true psychological problem becomes crystallized (the circle is closed), since it is critical reflection with oneself that creates an inert space, where the patient devotes his or her efforts to removing a psychic framework whose essential purpose is to block out confrontation with the genuine problem. In this context, the self-deceiver ends up losing the perspective of the original problem, and frequently appeals for help to escape from a disastrous circuit that was entered with the intention of calming the unpleasant perception of a conflict, but that will eventually leave the person without the capacity for response, or blind to this conflict, which, despite going unperceived, is nevertheless disturbing, and basically sustains and consolidates the psychological unease. In this regard, it is interesting to consider the example mentioned by Paul Watzlawick, recalling how the anthropologist Margaret Mead distinguished the Americans from the Russians. While the former, she observed, simulated headaches to elude responsibilities, the latter needed to actually suffer the headache for the same purpose (Watzlawick, 1975). So, perhaps a psychological problem is more than anything a "Russian headache", self-generated so as to tiptoe around the important aspects of life, and a headache that once it has struck, becomes more severe when one strives to find analgesics for a problem that the headache was only trying to get around.

In the end, self-deception, as we intend to represent it, coincides perfectly with the idea of the symptom as described in a recent essay (Pérez Álvarez, 2003), and overlaps with the expression or manifestation of a real problem, but also fulfils the function of an attempt to adapt, a truce or even a way of life.

THE STRANGE TRUTH OF PSYCHOLOGICAL TREATMENTS

On countless occasions, psychologists have to conceal things, keep quiet, tell half-truths, and make biased comments based on deception, lies or, at the very least, avoidance of the naked truth. Sometimes this is to avoid hurting the patient's sensibility, and in other cases it is merely a question of politeness. Nevertheless, we feel that pretence and appearances play a substantial role, rather than a superficial one, in the task of the therapist.

Using the example of medicine, we could say that the surgeon can operate without the patient's awareness. And medication functions relatively independently of the actual beliefs of the person taking it, but the same is not true in the context of psychology. If there is one thing that characterizes psychological therapy it is the crucial importance of the phenomenon of appearance, to the extent that it is impossible to carry it out without a "performance", without the psychologist "performing" for the patient and vice versa (it is even doubtful whether true therapy could take place without the awareness that the therapeutic process is actually happening). The doctor can be absent, but the psychologist has to be at least co-present.

In our view, a psychological treatment is somewhat similar to a game of football (similar analogies have been proposed previously: therapy appears as a game of chess, and in general as a game, and as a challenge full of unexpected turns, for example, in the novels of the existential psychologist Irvin D. Yalom [Yalom, 1992; 1996]). The game will determine victory, but in order to win, the game has to take place within a framework that imposes certain rules, but never guarantees success in advance. Psychological treatment is carried out in the framework of a ceremony (García Sierra, 2001), which we might call the psychotherapeutic ceremony. As regards the importance of the concept of ceremony for psychology, Juan Fuentes and Ernesto Quiroga have produced a significant work on the subject (Fuentes & Quiroga, 1998). What we are trying to point out is simply that therapy is always developed in the context of a series of transitory sequences that follow certain rules: sessions are more or less regulated in terms of time, the psychologist and the patient take turns to speak, corrections are made more in one direction than in the other, authority belongs more to one participant than to the other, and so on.

With this in mind, our position is as follows: a treatment or therapy is, above all, tactics put into practice, like the tactics employed in football by a coach, which have to continually be adjusted to the real conditions occurring on the field of play, or in the psychologists' consulting room (in this case, the "play·" is what is being said). Psychologists cannot simply apply a series of steps until they reach a goal, because they continually have to adjust

their steps in response to those of the patient, as a forward does in football when he faces an opposing defender, and this means –let us say it loud and clear– that therapy is a game of risky lies in which psychologists have to keep patients convinced that they will provide the solution their problem –a solution which (as we said above) is never assured in advance–, which will only be true insofar as psychologists can sustain during the process the lie that they possess that solution. All of this can be summed up euphemistically: psychologists, if they aspire to the name, have to maintain their credibility. But without recourse to euphemisms we would add: through diverse acts of sleight of hand.

Shaping is a procedure used by behaviour therapists that consists in starting out from a series of previous behaviours and gradually extending partial achievements until a final point or achievement is reached. An agoraphobic might, for example, accept going out in certain places but would not accept going out in others at all. The secret, we might say, consists in getting the patient to go out in places he would not accept by beginning with getting him to go out in those he accepts without much resistance. If we think carefully about this, we realize that what is really involved is the patient's will, and that to control it, it has to be in some way deceived, for in fact it is not clear that habituation might not be achieved by doing directly what the patient refuses to do, but it can be more effective to get him to do what he doesn't mind doing so as, eventually, to get him to do what he would never be persuaded to do, and which is what is really necessary for a successful outcome. Successive approach techniques are in this direction. It is not that a person with a phobia of lifts is incapable of going up to the sixth floor, but rather the psychologist has to get them to decide to go up to that floor, and moreover, it must be the patient him/herself who decides to go up in the lift voluntarily, despite having sought professional help because they are not prepared to go up of their own free will. It is not difficult to realize that all that comes in between contains a great deal of belief, more than of reality, since the client has to attribute to the procedure a value that is not strictly true: it is not the habituation that reduces the fear, and therefore permits the patient to go up in the lift, but rather the decision to go up that kickstarts the habituation process, and it is the patient's free decision that must always be the focal point of the psychologist's work.

Cognitive *rationalist* techniques also have their degree of deceitful skill, since they are often based on counteracting rigid catastrophic thinking with equally biased conflicting evidence, for the fact is that the therapist's discourse is frequently no more than a manner of speaking, which the therapist can actually readapt to each case to the extent of saying one thing or the opposite, as appropriate, with the functional aim of overcoming patients' rigidity, rather than of convincing them of another truth, which could be counter-productive.

Some techniques, such as *role interchange*, clearly reveal in what the therapeutic game consists, namely: finding the truth through pretence. The psychologist adopts the role of patient so that the latter can realize that some element of his/her discourse is an obstacle to progress. But why not tell the patient directly? The idea of the technique is that the client realizes without feeling offended or attacked, which can lead to defensive reactions or to the client ignoring the basics.

All therapies, it could be said, contain a good deal of paradoxical components, be they behavioural, cognitive or other types of therapy. But in reality, the above references to shaping, biases and role interchange should be interpreted as particular cases of a general form of approaching therapy. Within the psychological literature we find, in fact, an applied technique of a consistent, global nature, and which we believe is particularly successful if it fulfils the function for which it is designed and intended, namely, the technique of *paradoxical intention*. Since Adler, passing through Victor Frankl and up to the current Acceptance and Commitment Therapy, there is a tradition through which it has become a well established way of working. Within this psychological tradition, this technique can be interpreted not so much as a technique per se, or as a residual aspect of therapeutic programmes, but rather as an authentic way of dealing with psychological problems. And this is, moreover, the position to which we are committed in the present work.

Paradoxical intention clearly reveals the phenomenon of self-deception present in psychological disorders. A man with erectile dysfunction may desire sexual relations but not want to risk failure. A person with social phobia may desire relationships with other people but find it hard to accept the possibility of encountering setbacks in those relationships. A person may want to make advances to someone, but is afraid of turning bright red in the face. Someone who wants to slim may not fancy having to do exercise or go on a diet. In all of these cases

the patient focuses on the struggle against the secondary elements of a psychological nature (fear, anxiety, rumination, etc.), rather than confronting the original conflict. Thus, paradoxical intention resolves, or attempts to resolve, the problem with a disguise, that is, it tries to involve patients in paradoxical secondary elements in order to thrust them into a confrontation with the basic conflict. If as a result of anxiety a person begins to ruminate on how to avoid tripping over their own tongue, the therapist asks them to *want* to stumble over words, since in this way the ruminative element loses functional meaning, confronting the person with the conflict of speaking even at the risk of tripping up, which will quite probably increase their fluency of speech.

The manoeuvre of paradoxical intention is based on discrediting the secondary (psychological) conflict, trying to make the client become involved in provoking a problem that he or she attributes to an emotion or feeling, and not to the will to avoid a confronting a situation. And so, curious as it may seem, if patients bend their will to suffering the unpleasant psychological effects, these will disappear.

Paradoxical intention takes advantage of the self-deception of the person who experiences their problem (emotion as obstacle to the action of confrontation) to favour a psychological achievement (cognitive-emotional relief) through action. Success is clearly more than likely, since although the confrontation is set in motion with the intention of gaining psychological relief, the action is in fact dismantling the basic conflict that explains the entire framework of the problem. What paradoxical intention destroys is the excuse of putting the psychological content before the action, and this is achieved by making the patient think that through a paradoxical action, which will moreover show itself as effective, the adverse psychic content will disappear. And it probably will disappear, but, as we say, because it ceases to make any utilitarian sense for the person on actually confronting the primary conflict, that is, running the risk of taking a direction and not going round and round the roundabout indefinitely. It is equally important to mention the recommendation that paradoxical intention, given its importance in highlighting the base conflict, be presented (concealed) in the form of humour, through encouraging patients to laugh at themselves (Frankl, 1946)

Let us conclude by saying that the present work represents no more than a frank attempt to acknowledge that lying, strategy, the oblique approach, are essential and defining aspects of a large part of what we call Psychology, and that this should not necessarily give the discipline a bad name. Perhaps the same idea was expressed in another way by one of the fathers of our discipline, Alfred Binet, when in measuring intelligence he discovered that what he was really doing was assigning value to errors, and not to correct answers: "while Logic concerns itself with intellectual processes to do with the truth, Psychology is especially concerned with intellectual processes to do with error".

REFERENCES

Frankl, V. (1946/2004). *El hombre en busca de sentido*. Barcelona: Editorial Herder.

Fuentes Ortega, J.B. (1994) Introducción del "conflicto de normas irresuelto personalmente" como figura antropológica (específica) del campo psicológico. *Psicothema, 6* (3), 421-446.

Fuentes Ortega, J.B. & Quiroga Romero, E. (1998) A propósito de Mesmer y la Hipnosis. La distinción entre ceremonias abiertas y ceremonias cerradas. *Revista de Historia de la Psicología, 19* (2/3), 421-429.

García Sierra, P. (Ed.). (2001). *Diccionario filosófico "manual del materialismo filosófico"*. Oviedo: Pentalfa Ediciones.

Hayes, S.C. (1984). Making sense of spirituality. *Behaviorism, 12*, 99-110.

Luciano, M.C. & Hayes, S.C. (2001). Trastorno de evitación experiencial. *Revista Internacional de Psicología Clínica y de la Salud, 1* (1), 109-157.

Pérez Álvarez, M. (1999). Psicología clínica y iatrogénesis. In J. Buendía (Ed.). *Psicología clínica: Perspectivas actuales*. Madrid: Pirámide.

Pérez Álvarez, M. (2003). *Las cuatro causas de los trastornos psicológicos*. Madrid: Editorial Universitas.

Quiroga Romero, E. (1999). La personalidad como estructura dinámica conductual definida por la conjugación Autor-Actor. *Apuntes de Psicología, 17* (3), 309-325.

Watzlawick, P. (1975/1986). *El arte de amargarse la vida*. Barcelona: Editorial Heder.

Wilde, O. (1889/2000). *La decadencia de la mentira*. Madrid: Editorial Siruela.

Yalom, I. D. (1992/1999). *El día en que Nietzsche lloró*. Buenos Aires: Editorial Emecé.

Yalom, I. D. (1996/2000). *Desde el diván*. Buenos Aires: Editorial Emecé.

Papeles del Psicólogo, 2005. Vol. 26, pp. 115-128

PERSONALITY AND SOCIAL DESIRABILITY IN ORGANIZATIONAL SETTINGS: PRACTICAL IMPLICATIONS FOR WORK AND ORGANIZATIONAL PSYCHOLOGY

Jesús F. Salgado
University of Santiago de Compostela

This article consists of four sections. The first section presents a validity review of personality measures for predicting several organizational criteria, including job performance, training success, leadership emergence, leadership effectiveness, work accidents, job satisfaction, turnover, counterproductive behaviours, absenteeism, and salary. Secondly, the literature on social desirability and distortion of responses to personality questionnaires, one of the main problems of personality measures, is examined. This examination suggests that social desirability has an average effect size of 0.38 standard deviation units on the personality measures, inflating the scores in personnel selection settings. However, social desirability has no effect on the predictive validity of personality measures, and is not a moderator, mediator or suppressor variable. It is also observed that social desirability is not related to job performance. In the third section, the various strategies developed for reducing social desirability are reviewed, and only two are found to be effective: (a) warning applicants that their responses will be checked for social desirability and that distortion will be penalized, and (b) developing norms using job-applicant samples or samples including individuals responding to personality measures in contexts which can produce social desirability (e.g., promotion decisions). The article ends with some conclusions and some suggestions for practitioners in Work and Organizational Psychology.

Este artículo presenta una revisión de la validez de las diversas medidas de personalidad para predecir diversos criterios organizacionales, entre los que se incluyen, el desempeño laboral, el éxito en la formación, la emergencia del liderazgo, la eficacia del liderazgo, los accidentes laborales, la satisfacción laboral, la rotación en el empleo, las conductas contraproductivas, el absentismo y el salario. Seguidamente, se ocupa de examinar uno de los problemas principales a los que se han enfrentado las medidas de personalidad en el trabajo: la deseabilidad social y la distorsión de las respuestas a los cuestionarios. De este examen se desprende que la deseabilidad social tiene un efecto promedio de inflar (o desinflar) las puntuaciones alrededor de 0.38 unidades de desviación en contextos de selección, aunque no afecta a la validez predictiva de las medidas de personalidad, ni es una variable mediadora, moderadora o supresora de la validez de dichas medidas. También se observa que la deseabilidad social no está relacionada con el desempeño en el trabajo. En tercer lugar, se revisan las distintas estrategias utilizadas para reducir y neutralizar la deseabilidad y se observa que sólo dos de ellas son efectivas: (a) informar a los evaluados de que se examinarán sus respuestas en relación con la deseabilidad social y que la distorsión podrá tener consecuencias negativas para los distorsionadores, y (b) crear baremos a partir de muestras de solicitantes o de personas que contestan a las medidas de personalidad en contextos que puedan suscitar la deseabilidad social (p.e., decisiones de promoción). La última parte del artículo contiene las conclusiones y las sugerencias para los profesionales de la Psicología del Trabajo y las Organizaciones.

In the last fifteen years, the assessment of personality has become an issue of great interest for professionals and researchers in the field of Work and Organizational Psychology, given its utility for making decisions in connection with work-related processes. Different meta-analyses carried out in America, Europe, Africa and Asia have reached essentially the same conclusions: personality measures are good predictors of various relevant organizational criteria (see Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Hogan & Holland, 2003; Hough, 1992; Hurtz & Donovan, 2000; Mount & Barrick, 1995; Rothman, Meining & Barrick, 2002; Salgado, 1997; 1998; 2002; 2003; Yoo & Ming, 2002). The fundamental basis of these convergent conclusions is the use of the Five Factor Model (FFM) of personality as a taxonomy for integrating the results of hundreds of local validity studies carried out over more than 60 years. In accordance with this model, five broad dimensions of personality have been found to be replicable across different samples, in different cultures, for dif-

Correspondence: Jesús F. Salgado. *Departamento de Psicología Social, Básica y Metodología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.*
E-mail: psjesal@usc.es

TABLE 1
SUMMARY OF META-ANALYTICAL RESULTS ON THE RELATIONSHIP BETWEEN PERSONALITY MEASURES AND VARIOUS ORGANIZATIONAL CRITERIA AND VARIABLES
(Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance[a]			
Conscientiousness	133	33,668	.33
Emotional Stability	108	19,880	.21
Extraversion	111	21,916	.10
Openness to Experience	82	13,895	.09
Agreeableness	110	21,911	.19
Training Proficiency[b]			
Conscientiousness	20	3,909	.31
Emotional Stability	25	3,753	.09
Extraversion	21	3,484	.28
Openness to Experience	18	3,177	.33
Agreeableness	24	4,100	.14
Leadership Emergence[c]			
Conscientiousness	17	n/a	.33
Emotional Stability	30	n/a	.24
Extraversion	37	n/a	.33
Openness to Experience	20	n/a	.24
Agreeableness	23	n/a	.05
Effective Leadership[c]			
Conscientiousness	18	n/a	.16
Emotional Stability	18	n/a	.22
Extraversion	23	n/a	.24
Openness to Experience	17	n/a	.24
Agreeableness	19	n/a	.16
Job Satisfaction[d]			
Conscientiousness	79	21,719	.26
Emotional Stability	92	24,527	.29
Extraversion	75	20,184	.25
Openness to Experience	50	15,196	.02
Agreeableness	38	11,856	.17
Counterproductive behaviours (Validity reversed)[e]			
Conscientiousness	13	6,276	.26
Emotional Stability	15	3,107	.06
Extraversion	12	2,383	-.01
Openness to Experience	8	1,421	-.14
Agreeableness	9	1,299	.20
Turnover (Validity reversed)[e]			
Conscientiousness	5	748	.31
Emotional Stability	4	554	.35
Extraversion	4	554	.20
Openness to Experience	4	554	.14
Agreeableness	4	554	.22
Accidents at work (Validity reversed)[f]			
Conscientiousness	9	1125	.30
Emotional Stability	13	1198	.28
Extraversion	12	1524	-.09
Openness to Experience	7	570	-.50
Agreeableness	7	420	.61

Note: K= number of studies; N= Total sample size; n/a= not available; [a] Salgado (2004); [b] Barrick, Mount & Judge (2001); [c] Judge, Bono, Ilies & Gerhardt (2002); [d] Judge, Heller & Mount (2002); [e] Salgado (2002); [f] Clarke & Robertson (2005).

ferent languages and with different assessment techniques. The names of these factors vary among the different researchers in the field of the psychology of personality in the work context, but the labels most widely used are those suggested by Costa and McCrae (1992). For them, the five personality dimensions or factors would be: emotional stability (versus neuroticism), extraversion (versus introversion), openness to experience (versus closure to experience), agreeableness (versus antagonism) and conscientiousness (versus irresponsibility). The names of the emotional stability and extraversion factors are those on which there is greatest consensus. The other three, perhaps because they are more recent, are denoted by a variety of terms. For example, the openness to experience dimension has also been called culture or intellectuality; the agreeableness dimension has been called friendliness, and the conscientiousness dimension has been labelled dependability, prudence, responsibility or need for achievement. The results of the meta-analyses mentioned above have shown that two personality factors, emotional stability and conscientiousness, are valid predictors of job performance in all occupations. It has also been shown that these two factors, together with extraversion, are valid predictors of training proficiency, and that the conscientiousness and agreeableness factors predict counterproductive behaviour.

In addition to the above, research on personality in organizational settings has demonstrated that personality measures focusing on occupational criteria (*Criterion-focused Occupational Personality Scales*, COPS), such as integrity tests, client orientation scales, management potential scales, stress tolerance scales or commercial potential scales are excellent predictors of diverse organizational criteria, including job performance, training proficiency and counterproductive behaviours (see Ones & Viswesvaran, 2001a and b, for a summary). Research has also shown that these measures (COPS) are a combination of three of the basic personality factors: emotional stability, agreeableness and conscientiousness. Tables 1, 2 and 3 provide a summary of various personality dimensions and composites for the prediction of different organizational criteria.

In the light of the results of the meta-analytical studies mentioned, it is clear why personality measures have been so extensively used in organizational decisions in the last decade, and why they have aroused the interest of professionals. Although they have primarily been used

for purposes of personnel selection, they have also been employed in training processes and personnel development (e.g., *coaching*), and for establishing competence profiles.

Nevertheless, and despite this recent success of personality measures for predicting job performance, their use in organizational contexts is not without its problems and difficulties. Among these problems, distortion and response bias are two of those that have received most interest, and about which there has been most concern; consequently, there has been an abundance of studies dealing with these issues in recent years – though it is in fact sixty years since Mehl and Hathaway (1946) and Ellis (1946) demonstrated that people, when taught to do so, can distort responses to personality questionnaires.

DISTORTION OF RESPONSES TO ITEMS OF PERSONALITY MEASURES

Distortion and response bias to personality measures items, especially in organizational contexts, where decisions based on responses to personality questionnaires have important implications for the respondents (e.g., being hired or not), have received considerable attention from research over the last 50 years or more. Response distortion can be either positive (e.g., trying to make a good impression) or negative (e.g., trying to suggest a psychological problem or disorder). In the context of organizations, although the latter type of bias exists in certain circumstances, it is the former type, positive distortion, that has given the most cause for concern, given that personality measures have been used primarily for purposes of personal selection, so that 'trying to give

TABLE 2
SUMMARY OF META-ANALYTICAL RESULTS ON THE RELATIONSHIP BETWEEN COPS AND VARIOUS ORGANIZATIONAL CRITERIA AND VARIABLES
(Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance			
Integrity (personality tests)[a]	102	27,081	.37
Drugs and Alcohol Scales[b]	7	1,436	.19
Stress Tolerance Scales[b]	13	1,010	.42
Client Orientation Scales[b]	33	6,944	.39
Violence Scales[c]	14	4,003	.41
Training Proficiency			
Integrity (personality tests)[d]		2,364	.38
Counterproductive behaviours (Validity inverted)[e]			
Integrity (personality tests)[a]	138	158,065	.32
Stress Tolerance Scales[b]	5	594	.42
Client Orientation Scales[b]	5	740	.42
Violence Scales[c]	4	533	.46
Absenteeism			
Integrity (personality tests)[e]	16	5,435	.36

Note: K= number of studies; N= Total sample size; [a] =Ones, Viswesvaran & Schmidt (1993); [b] = Ones & Viswesvaran (2001a); [c] = Ones & Viswesvaran (2001b); [d]= Ones & Viswesvaran (1998a); [e]= Ones, Viswesvaran & Schmidt (2003).

TABLE 3
SUMMARY OF THE META-ANALYTICAL RESULTS BETWEEN SEVERAL PERSONALITY VARIABLES (AND MODELS) AND SEVERAL ORGANIZATIONAL CRITERIA AND VARIABLES.
(Original source: Salgado & De Fruyt, 2005)

Dimension	K	N	Validity
Job Performance			
Conscientiousness-NFFM[a]	36	5,874	.18
Emotional Stability-NFFM[a]	25	4,541	.05
Extraversion-NFFM[a]	26	4,338	.08
Openness to Experience-NFFM[a]	29	4,364	.08
Agreeableness-NFFM[a]	31	4,573	.13
Generalized Self-Efficacy[b]	11	1,506	.43
Locus of Control[c]	35	4,310	.22
Self-Esteem[c]	40	5,145	.26
Training Proficiency			
Generalized Self-Efficacy[b]	4	422	.29
Job Performance			
Emotional Intelligence[d]	19	2,652	.24
Job Satisfaction			
Positive Affectivity[e]	15	3,326	.49
Negative Affectivity[e]	27	6,233	-.33
Generalized Self-Efficacy[c]	8	1,411	.29
Locus of Control[c]	80	18,491	.32
Self-Esteem[c]	56	20,819	.26
Salary			
Generalized Self-Efficacy[b]	5	468	.28
Absenteeism			
Generalized Self-Efficacy[b]	4	718	.21

Note: K= number of studies; N= Total sample size; NFFM= measures from questionnaires not based on the Five Factor Model; [a] =Salgado (2003); [b]=Salgado & Moscoso (2000); [c] =Judge & Bono (2001) [d]=Van Rooy & Viswesvaran (2004); [e] =Connolly & Viswesvaran (2000).

a good impression' is frequently an option as a response to tests and questionnaires. In view of the fact that response distortion in a favourable direction can favour candidates' possibilities of being hired, Seisdedos (1988), in one of the few studies carried out in Spain on this phenomenon in organizations, has called this bias "Intelligent adaptation". However, it is typically labelled as 'Social desirability', 'Sincerity' or 'Motivational Distortion' – though other authors refer to it as 'Faking', 'Infrequent Virtues' or 'Response Distortion'. Of all these terms, 'Social desirability' is that most commonly used, and covers all the others.

Social desirability (or response distortion) has been defined as "the tendency to endorse items in response to social or normative pressures instead of providing veridical self-reports" (Ellingson, Smith & Sackett, 2001, p.122). Although it has often been seen as a unidimensional construct, and many measures provide only a global measure of it, according to the most recent research, mainly that of Paulhus (1984, 2002), desirability has two dimensions, which have been called *impression management* and *self-deception*. Impression management indicates a tendency to intentionally adapt one's public image with the aim of being favourably viewed by others. Self-deception, on the other hand, refers to the unintentional tendency to describe oneself in a favourable manner, and is expressed through self-descriptions that are positively biased but in which one honestly believes. In this sense, impression management is a voluntary manipulation of one's own image so that others perceive us in a positive way, whilst self-deception is not a deliberate manipulation, though it may lead to distortions in others' perceptions of us. Taking into account this distinction between the two dimensions of social desirability, in the context of the Psychology of Work and Organizations, the effects of impression management on personality measures scores appear to be the more relevant.

The distinction between impression management and self-deception is relevant in the light of the suggestion by some researchers in the psychology of personality that there are individual differences in social desirability (Block 1965; McCrae & Costa, 1983). This means that social desirability may be not simply a tendency to adapt to situations, but rather a stable personality characteristic that would indicate more substantive and significant differences. For example, McCrae and Costa, (1983), Ones, Viswesvaran and Reiss (1996), and Salgado, Iglesias and Remeseiro (1996) have found that social desirability correlates with emotional stability and with conscientiousness. As McCrae and Costa (1983) point out, this would mean that a person who genuinely scores highly in conscientiousness, and who is emotionally stable and cooperative (scoring highly in agreeableness) will also score highly in social desirability. However – and this is the paradox –, this person may be honest and reliable, but would be "guilty" of distortion or lying in personality questionnaires.

The effects of social desirability on personality measures and on other assessment instruments (e.g., interviews) are widely documented in the literature, and are particularly well known to those responsible for personnel selection in organizations. The principal effect is that social desirability tends to inflate (increase) scores in the dimensions that are (or that candidates believe to be) positively related to job performance and to deflate (reduce) scores in those dimensions that are (or that candidates believe to be) negatively related to job performance. Recent research has reviewed meta-analytically, and through studies with large samples, the effects of social desirability on responses to personality measures. In general, such empirical research has consisted in the use of three types of design: (a) comparisons between groups in laboratory situations, with participants instructed for distorting their responses; (b) within-subject comparisons in laboratory situations, with participants instructed to distort, and (c) comparisons in real selection situations, examining the differences between those who show social desirability and those who do not.

Ones and colleagues (Ones & Viswesvaran 1998a and b; Ones, Viswesvaran & Reiss, 1996; Viswesvaran, Ones & Hough, 2001), Hough and colleagues (1998; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Hough & Paullin, 1994) and Christiansen and colleagues (Christiansen, Goffin, Johnston & Rothstein, 1994; Goffin & Christiansen, 2003) are those that have done most research on the effects of social desirability in organizational settings, though other researchers have made relevant contributions. The main results are as follows. In the studies with comparisons between groups in laboratory situations with induced positive distortion it was found that, on questionnaires measuring personality dimensions (e.g., the *Big Five*), "fakers" score, on average, 0.6 standard deviation units more than "non-fakers". In terms of T scores (mean=50; SD=10), this means that the fakers group would score an average of 56, as against a

score of 50 for the non-fakers group. When studies make within-subject comparisons in laboratory situations with induced distortion and honest response, the difference in the scores on personality dimensions measures between the two conditions for the same participants is 0.72 standard deviation units, which in terms of T scores means that the distortion increases a person's score from 50 to 57.2 points. When type of design is not taken into account, the difference is 0.5 standard deviation units. Thus, these results indicate that social desirability has effects on distorters' scores of slightly over half a standard deviation unit. They also show that type of research design has important effects on the size of the distortion. This same type of study was carried out with composite personality measures (COPS), such as integrity tests, with quite similar results, since the difference between people instructed to respond in a positively distorted manner and those instructed to respond honestly was 0.50 standard deviation units.

The results described above can be considered as indicative of the effects of social desirability in situations of maximum performance (when the aim is to find the greatest possible difference between the conditions of social desirability and honesty). However, studies on the effects of social desirability in "real" conditions, that is, which compare the responses of fakers and non-fakers in genuine selection situations, would reflect the typical performance situation (where the aim is to find the usual difference between the two mentioned conditions in a typical work context). In this regard, Hough (1998) carried out three particularly relevant studies. In the first of these she compared a sample of 963 telecommunications employees with a sample of 14,442 job applicants in the same industry, finding in a measure of responsibility (one of the sub-dimensions of the conscientiousness factor) an average distortion of 0.45 standard deviation units on the side of the applicants. In a second study, in relation to positions in the local police, with a sample of 508 employees and 24,433 applicants, using a measure of conscientiousness, Hough found the average distortion to be 0.33 standard deviation units on the side of the applicants. A third study, with a sample of 270 national guards and 681 applicants, found an average distortion of 0.13 on a personality scale composed of the experience-seeking and self-esteem dimensions. Other studies carried out by Hough, Eaton, Dunnette, Kamp and McCloy (1990) produced similar results. Therefore, on the whole, the available data indicate that in the third type of research design (comparison of employees and applicants in real situations), the average effects of social desirability, while they exist, are much smaller than those found in laboratory settings. In other words, in situations of typical performance (e.g., personnel selection), the effects of social desirability are less than those found in situations of maximum performance (e.g., experimental manipulation).

In addition to the effect of inflating scores on personality measures, a second possible effect often mentioned, and which is probably that which causes most concern among Work and Organizational Psychology professionals using personality measures for personnel management (hiring decisions, training, promotion, etc.), is that associated with the impact of social desirability on the validity of such measures. Some researchers, and many professionals, believe that social desirability has a negative effect on the validity of the measures, reducing it, and thus making them invalid. In other words, the validity of personality measures for predicting job performance and other organizational criteria would be substantially reduced, or even cancelled out, by the effects of desirability (Goffin & Christiansen, 2003; Mueller-Hanson, Heggestad & Thornton, 2003; Rosse, Stecher, Miller & Levin, 1998). This belief has led to serious doubts over the use of personality measures in organizational settings, and even to some professionals openly rejecting their utility for decision-making.

In relation to this, meta-analytical and individual studies with large samples (N > 1200) have recently been carried out to explore the effects of social desirability on the validity of personality measures. Thus, for example, with regard to construct validity, Ones, Viswesvaran and Reiss (1996) showed that social desirability has scarcely any effect on the convergent and discriminant validity of personality measures, the average increase in correlations being of the order of .015, that is, irrelevant in practice. As far as criterion (predictive) validity is concerned, Ones and Viswesvaran (1998b) have examined the effects of social desirability (whether it be considered as a moderator, mediator or suppressor variable). The results of their meta-analyses indicate that social desirability has no impact on the validity of personality measures, whether these refer to the basic dimensions (e.g., the *Big Five*) or personality composites (e.g., integrity), with validity remaining essentially the same once the effects of social desirability have been accounted for. The studies by Hough (1998; Hough et al., 1990) obtained

results that were basically similar. Such work has demonstrated the erroneous nature of the beliefs of some researchers and many professionals about the negative effects of social desirability on the validity of personality measures.

HOW IS SOCIAL DESIRABILITY DETECTED (MEASURED)?

Given the interest generated by social desirability and its effects on personality measures scores, several researchers have considered the question of how to detect it; consequently, over the years, a series of instruments have been designed, though almost all of them take as their source the *Minnesota Multiphasic Personality Inventory* (MMPI). The develepors of the MMPI, Hathaway and McKinley, actually designed two scales for detecting possible distortion in responses to the questionnaire. One is the so-called K scale, for detecting negative bias, or the tendency to present a poorer image of oneself; the other is the *Lie* (L) scale, sometimes described as the Sincerity scale, and which is aimed at detecting positive bias, or the tendency to project a good image of oneself. Both scales were developed as criterion-focused tests, as the rest of the MMPI scales had been. Following the line of the MMPI, Harrison Gough, creator of the *California Personality Inventory* (CPI; 1987) and a student of Hathaway, also developed a scale for measuring social desirability in his questionnaire, calling it the *Good Impression Scale*. Given that the CPI was designed to assess normal (adjusted) personality, in contrast to the MMPI, which was intended for the assessment of personality disorders, Gough was most interested in the tendency to distort positively. Eysenck, in his first personality questionnaire, the *Maudsley Personality Inventory* (MPI; predecessor of the *Eysenck Personality Inventory*, EPI), also included a scale for assessing social desirability. This scale was also derived from the MMPI's L scale. Likewise, many later researchers have created scales for measuring social desirability, and a good deal of personality questionnaires include among their items some scale or other measure in relation to such distortion. In addition to those mentioned above, other popular instruments include Edwards' social desirability scale (1957), Crowne-Marlowe's social desirability scale (1964), Eysenck's sincerity scale (Eysenck & Eysenck, 1964; originally a lying scale), the positive motivational distortion scale of the 16PF (Cattell, Eber & Tatsuoka, 1970) or the social desirability scale of the Occupational Personality Questionnaire (SHL, 1999). All of these scales were designed on the basis of social desirability being a unidimensional concept. However, since the research by Paulhus referred to above, it has been accepted that social desirability comprises two dimensions, and the *Balanced Inventory of Desirable Responding* (BIDR, Paulhus 1984, 2002) is the most popular of the modern inventories. This inventory consists of 40 items, with two 20-item subscales, which assess 'impression management' and 'self-deception'.

Nevertheless, it should be pointed out that many recent personality questionnaires, especially those based on the Five Factor model, do not use a scale for detecting social desirability. This is the case, for example, of the NEO-PI-R (Costa & McCrae, 1992), the *Hogan Personality Inventory* (HPI, Hogan & Hogan, 1995), the IP/5F (Salgado, 1996) or the *Jackson Personality Inventory* (JPI, Jackson, 1994).

STRATEGIES FOR REDUCING THE EFFECTS OF SOCIAL DESIRABILITY

Having examined the effects of social desirability in organizational contexts, and seeing that they involve the inflation or deflation of scores on measures of some dimensions and facets of personality, the following question concerns how these effects can be reduced or cancelled out.

In this regard, over the years a variety of strategies have been proposed. Among them are the following: (a) use of scales with forced-choice items, the items having been paired according to their similarity in social desirability (this mode is also called 'ipsative measure' or 'ipsative score'); (b) use of scales for detecting social desirability and discarding of respondents who score moderately high (e.g., 2 standard deviation units over the mean) on these scales; (c) use of detection scales and subjective "adjustment" of the personality-measure scores of respondents scoring moderately high on the social desirability scales; (d) use of detection scales and mechanical "correction" of the personality-measure scores of respondents who distort moderately highly, and use of a mathematical formula for this purpose; (e) warning respondents about the existence of detection methods in the assessment, and instructing them about the possible consequences of distortion; (f) developing specific norms for samples of applicants, rather than using the norms of normative samples from the general population. Table 4 provides a summary of these strategies.

Currently, there is very little use of scales with forced-choice format for assessing personality in organizational

contexts, especially for purposes of personnel selection. Nevertheless, there are some questionnaires that use this format, with the intention of reducing distortion in responses. Examples of personality questionnaires that employ this approach are the *Occupational Personality Questionnaire* 3.2 (OPQ 3.2; SHL, 1999), the Thomas *Personality Inventory*, also known as the DISC (Thomas International), or the *Description en Cinq Dimensions* (D5D, Rolland & Mogenet, 2001). The basic assumption of those who design this type of questionnaire is that if the items are grouped, for example, in fours, with similar social desirability, and respondents have to indicate which of the items best defines them and which is the least appropriate (discarding the other two options), then the final response will better reflect their personality characteristics and will eliminate the effects of social desirability. Hicks (1970, p.181) suggested that to justify the use of forced-choice (ipsative) measures, three conditions were required: (1) that there is a marked bias in responses to personality questionnaires, (2) that this bias reduces the validity, and (3) that the forced-choice format reduces the bias and increases the validity to a greater extent than other, non-ipsative controls of bias. Hicks concluded that no case had occurred in which these three conditions were jointly met. Twenty-five years later, Bartram (1996) considered that Hicks' conclusion remained true. In this regard, the following should be noted: (a) the most recent and exhaustive research, mentioned here in previous sections, has demonstrated that there is a bias, and that it can be important in real assessment situations in organizational contexts (e.g., personnel selection); and (b) research has also shown that social desirability does not reduce validity. As regards the third condition, referring to the fact that forced-choice (ipsative) formats reduce bias and increase validity, various studies have dealing with this issue have appeared in recent years. For example, Christiansen, Burns and Montgomery (2005) carried out a series of studies showing that (1) forced-choice scales are as susceptible to distortion as traditional normative scales, and (2) respondents with higher scores in measures of general mental ability are more successful in improving their scores on forced-choice scales (in the sense of more closely fitting the desired profile), suggesting that such scales are manipulable according to respondents' general mental ability. Baron (1996), a staunch defender of this format for personality measures, acknowledges that a small group of candidates powerfully distort scores on these questionnaires. Consequently, Hicks' third condition, in the light of recent data, is not fulfilled in practice. Thus, and bearing in mind that this response format has many and serious limitations of a psychometric nature, with regard to its reliability, its factorization, its validity and comparisons between individuals (see, for example, the reviews by Baron, 1996; Bartram, 1996; Closs, 1996; Cornwell & Dunlap, 1994; Hicks, 1970; and Meade, 2004), the decision on this strategy is that it should be discarded, and not used for purposes of assessment in organizational contexts involving the comparison of persons among one another (e.g., selection).

The second strategy, discarding those respondents who score highly on social desirability scales, introduces serious complications of a theoretical, practical and possibly even legal nature. From the theoretical point of view, it is possible, as McCrae and Costa (1983) point out, that persons who are totally honest in their responses, but with certain personality characteristics that fit the typical profile of the distorter, would be excluded without further consideration. This would be a clear case of 'false positives', which would be rejected as a result of this strategy. Furthermore, and also from the theoretical perspective, it would be necessary to demonstrate that those scoring highest on motivational distortion scales subsequently present a job performance inferior to that of non-distorters. And not only has this not been demonstrated, but it has also been shown that there is no relationship between social desirability and job performance (Ones & Viswesvaran, 1998b). Nor has it been demonstrated that respondents who distort, and consequently present higher scores in some personality dimensions (e.g., conscientiousness) related to job performance, later show (after being hired) poorer job performance than those who obtain similar scores without distorting their responses. From the practical point of view, the exclusion of a number of persons from the set of respondents may be counter-indicated if the selection ratio is high (e.g., it approaches 1) – that is, if the numbers of vacancies and of candidates are similar. Finally, from the legal perspective, the exclusion of candidates on the basis of their distorted responses cannot be defended (in court, for instance) when it is known, as is now the case, that there is no negative relationship between social desirability and job performance. Therefore, having distorted cannot legally be a reason for exclusion, and any candidate lodging an appeal against a decision based on this point would have a very good chance of a favourable verdict.

Consequently, this strategy should not be used by professionals from the field of Work and Organizational Psychology in certain tasks carried out in labour-related contexts.

The third and fourth strategies consist in using detection scales and subjectively "adjusting" the scores on the personality measures of respondents who score moderately highly on the social desirability scales, or mechanically "correcting" the personality-measure scores of moderately high distorters, using a mathematical formula. As it can be appreciated, the two strategies are quite alike, and involve similar problems. Correction of scores, increasing or reducing them according to the degree of distortion detected on the social desirability scale, has been a widely used strategy among psychologists in general and those from the field of work and organizations in particular. This is due to the fact that two of the most popular personality questionnaires – MMPI and the 16PF – include systems for 'correcting' the scores. In the case of these two questionnaires, the strategy is the fourth one, that is, mechanical correction (based on a regression equation derived from motivational distortion), but it has led to many professionals taking a subjective approach to correction, based on their experience and the evaluator's 'clinical eye', and to their adjusting the scores accordingly. A problem common to the two methods, and usually overlooked by professionals, is that correction of the scores modifies the construct validity of the question-

TABLE 4
STRATEGIES FOR THE REDUCTION OF SOCIAL DESIRABILITY AND SUGGESTIONS FOR PROFESSIONALS

Strategy	Description	Limitations	Effectiveness	Recommendation
Forced-Choice Scales	Choose between items with similar degree of social desirability	Theoretical, methodological and practical	Limited	Not recomended
Use Social Desirability scales and discard candidates scoring highly on them	Include a measure of distortion and exclude those who score above a cut-off point	Theoretical, practical and legal	Ineffective	Not recomended
Use Social Desirability scales and adjust distorters' scores (subjective strategy)	Scores of candidates considered to be fakers are adjusted "subjectively", based on the assessor's experience	Theoretical and practical	Ineffective	Not recomended
Use Social Desirability scales and adjust distorters' scores (objective strategy)	Scores of candidates considered to be fakers are adjusted "subjectively", based, for example, on a regression equation	Theoretical and practical	Ineffective	Not recomended
Warn candidates	Candidates are warned of the possibility of being eliminated or penalized if they distort their responses	None	Effective	Recommended
Develop norms based on samples of job applicants	Calibrate candidates' scores after creating a norm developed with samples of job applicants, rather than a sample of the general population	None	Effective	Recommended

naires (see Ones & Viswesvaran, 1998a and b), which means that the measure and its reliability are modified, without improvement to its criterion (predictive) validity, which may indeed be adversely affected (see, Hough 1998). Thus, scores derived from the correction and adjustment of raw data may fail to correspond to the respondent's actual personality characteristics. Moreover, the third strategy is practically unfeasible when dealing with a large number of candidates, since it requires the examination of each particular profile in order to carry out the correction. Consequently, these two strategies are also at odds with sound professional practice based on current empirical knowledge. It is not surprising, therefore, that in the latest version of the 16PF the correction of scores is no longer used.

The fifth strategy referred to above consists in warning respondents to the personality questionnaires that these incorporate methods for detecting social desirability and other possible biases, and asking them to be as sincere and honest as possible. At the same time, they are warned that those respondents detected as fakers may be disregarded as candidates for further consideration, or penalized in some other way appropriate to the case. For example, Hough (1998) used this strategy in several selection processes, warning applicants that those who distorted their scores on the personality questionnaire would be detected, and that those who were identified as having provided exaggeratedly favourable self-descriptions would be disqualified from the selection process for six months, after which time they could reapply for assessment. Although Hough did not assess the effectiveness of this fifth strategy (being more interested in comparing the difference in bias between employees and applicants), from the data she provides (Hough, 1998) and those of Ones, Viswesvaran and Reiss (1996), it is possible to estimate its effectiveness. Hough's three studies (1998) have an accumulated total sample of 40,297 persons, and average distortion weighted by number of applicants in relation to number of employees is 0.37 standard deviation units. Bearing in mind that the present case involves comparing fakers and non-fakers, we can use as an estimation of this comparison that obtained by Ones et al. (1996) for comparisons between groups in situations of maximum performance, which was 0.60 standard deviation units. Thus, deducting 0.37 standard deviation units corresponding to fakers in selection situations from the figure of 0.60 (the maximum distortion that can be obtained), the resulting value is 0.23 standard deviation units less, which is directly attributable to the strategy of warning the candidates. The results of these two studies, then, indicate that this strategy is effective in the reduction of social desirability, bringing it down by more than 38%. Moreover, given that the penalization suggested by Hough merely postpones the decision on hiring distorter candidates, this strategy is legally defensible. More recently, McFarland (2003) examined, in a laboratory setting, the effects of this strategy for reducing distortion, and assessed its effects on applicants' reactions in relation to the perceived organizational justice. McFarland's results indicate that the strategy was highly effective in reducing social desirability (0.45 standard deviation units on average), and that, moreover, it had no negative effects on applicants' perceptions (in one case these were even positive) and reduced multicollinearity between the personality variables. Consequently, this strategy is both valid (it reduces distortion) and economical, as there is no need to actually include a measure of social desirability – it is sufficient to inform candidates that there is one and that it can lead to their being penalized.

The final strategy mentioned consists in developing specific norms for contexts of personnel selection. In other words, instead of using the norms corresponding to the general population or to particular groups developed on the basis of scores obtained in situations where respondents have no interest in distorting, this strategy involves developing norms from scores obtained in situations where people have a direct interest in distorting (e.g., in personnel selection situations). It is obvious that the norms drawn up in this way will already include a part of the normative score corresponding to distortion, and which is common to all respondents. Thus, the social desirability will have already been partially neutralized. That is, a framework will have been developed and applied that is common to all respondents, rather than using a framework that is only common to those who never distort, such as one based on norms created with a normative population that responded to the questionnaire in a situation of null social desirability (e.g., with samples of students not induced to distort).

CONCLUSIONS AND SUGGESTIONS FOR PRACTICE

The possibility of responses to personality questionnaires being distorted, either positively or negatively, is a real one, and the phenomenon, which has been known of for more than sixty years, continues to be of great concern to

professionals in the field of Work and Organizational Psychology with responsibilities in the use of personality measures in their daily work (e.g., personnel selection). Such concern has led to different approaches to the use of personality measures among these professionals, the three most common being: (a) given the possibility of distortion in responses, to reject the use of personality measures, considering that such distortion invalidates their predictive capacity, so that appropriate decisions cannot be made on the basis of these instruments. The answer in this case has been to remove them from the toolbox of the organizational psychologist; (b) despite acknowledging the potential for distortion, to consider that personality measures continue to be valid and useful for professional work, and to seek formulas for overcoming or neutralizing this deficiency. This is the case of psychologists who have used corrective measures for the scores; (c) to consider that social desirability is not a significant problem, and that it would be much worse to go without the use of personality measures in one's professional practice.

Studies carried out in the last 20 years, and especially in the last ten years, have permitted researchers to reach sound conclusions on the effects of social desirability and possible ways of offsetting it. These conclusions can be summarized as follows:

1) Personality measures – those based on the Big Five factor model, criterion-focused occupational personality questionnaires, and instruments based on alternative models – are excellent predictors of job performance, training proficiency, counterproductive behaviours, leadership, job satisfaction, organizational commitment, knowledge acquisition, and many other relevant criteria for organizations. In some cases, personality measures are the best predictors of such criteria (e.g., counterproductive behaviours or job satisfaction).
2) Social desirability affects all personality assessment methods based on questionnaires, including those designed to be free of its effects, such as forced-choice questionnaires or ipsative measures. That is, no self-report personality measure is immune from the effects of social desirability.
3) Social desirability, as a relevant form of distortion, affects only a small percentage of those assessed in organizational processes.
4) The effect of social desirability varies depending on the way such desirability is triggered. In typical selection situations, desirability has an average effect of 0.38 standard deviation units. In terms of T scores, this means an increase or decrease in scores of 3.8 points.
5) Social desirability comprises two factors, impression management and self-deception. In organizational contexts the former is the more relevant, and that which inflates (or deflates) scores on personality questionnaires.
6) Social desirability is not related to job performance, and therefore does not affect the validity of personality measures. In other words, if the effects of social desirability are eliminated from personality measures, these do not have higher validity for predicting job performance. Therefore, the belief mentioned above, referring to the notion that social desirability invalidates the predictive capacity of personality measures, is simply erroneous.
7) Over the years, diverse strategies have been developed for neutralizing or reducing the effects of social desirability, and the majority have turned out to be ineffective or ill-advised:
   a) The strategy of forced-choice items does not prevent them from being distorted, does not improve the predictive validity of the measures, negatively affects their reliability, involves serious psychometric problems, and does not permit comparisons between respondents, so that it cannot be used appropriately in those situations that require the comparison of candidates (e.g., selection, training). Therefore, it should not be used.
   b) The strategy of discarding candidates who score highly on the social desirability scales has serious theoretical, practical and legal complications. Therefore, it should not be used.
   c) The strategy of "correcting" scores in a subjective manner has theoretical and practical complications. It affects the construct and predictive validity of the instruments, and its use is unfeasible with large numbers of candidates. Therefore, it should not be used.
   d) The strategy of "correcting" scores mechanically (e.g., by means of a regression equation) involves the same theoretical complications as the previous strategy. Therefore, it should not be used.
   e) The strategy of warning candidates about the existence of detectors of distortion and alerting

them to the potential consequences of such distortion considerably reduces distortion, and is an economical means of controlling social desirability. It should be used in all processes and assessments in which social desirability may be present.

f) The strategy of creating norms based on samples of job applicants neutralizes the effects of social desirability. Therefore, it should be used in all selection processes. And in these cases professionals should use instruments that provide such norms.

8) The combination of strategies of warning the candidates and using norms developed with samples of job applicants produces optimum results for the reduction of social desirability. This is the best option for professionals.

The above conclusions are based on the evidence currently available, chiefly obtained from meta-analyses and studies with large samples (N > 1200). In the light of this evidence, professionals in the field of Work and Organizational Psychology can confidently use personality measures for making decisions, with the limitations inherent to any psychological measure (i.e., reliability, validity and utility), without considering social desirability as a problem that invalidates them. Indeed, it seems that all the fuss over social desirability has resulted from the artificial creation of a problem with scarce relevance for the profession.

REFERENCES

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*, 49-56.

Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Barrick, M.R., Mount, M.K. & Judge, T. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9-30.

Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology, 69*, 25-39.

Block, J. (1965). The challenger of response sets: *Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century Crofts.

Cattell, R.B., Eber, D. & Tatsuoka, M. (1970). *Handbook of 16PF*. Champaign, IL: IPAT.

Christiansen, N.D., Burns, G.N. & Montgomery, G.E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*, 267-307.

Christiansen, N.D., Goffin, R.D., Johnston, N.G. & Rothstein, M.G. (1994). Correcting the Sixteen Personality Factors test for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847-860.

Clarke, S. & Robertson, I.T. (2005). A meta-analytic review of the Big Five personality factors and accidents involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78*, 355-36.

Closs, S.J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology, 69*, 41-47.

Cornwell, J.M. & Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Wilson (1991. *Journal of Occupational and Organizational Psychology, 67*, 89-100.

Costa, P. T. & McCrae, R. R. (1992). *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

Crowne, D.P. & Marlowe, D. (1964). *The approval motive*. New York: Wiley.

Connolly, J.J. & Viswesvaran, C. (2000). The role of affectivity in job satisfaction: A meta-analysis. *Personality and Individual Differences, 29*, 265-281

Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.

Ellington, J., Sackett, P.R. & Hough, L.M. (1999). Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.

Ellington, J.E., Smith, D.B., & Sackett, P.R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*, 122-133.

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin, 43*, 385-440.

Eysenck, H.J. & Eysenck, S.B. (1964). *The Eysenck Personality Inventory manual*. London: Routledge.

Goffin, R.D. & Christiansen, N.D. (2003). Correcting personality tests for faking: A review of popular personality tests and initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340-344.

Gough, H.G. (1987). *Manual: The California Personality Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.

Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100-112.

Hough, L.M. (1992). The "Big Five" personality variable construct confusion: Description versus prediction. *Human Performance, 5*, 139-155.

Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*, 209-244.

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. & McKloy, R.A. (1990). Criterion-related validities of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

Hough, L.M. & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G.S. Stokes, M.D. Munford & W.A. Owens (Eds). *The biodata handbook: Theory, research and application*. Palo Alto, CA; Consulting Psychologists Press.

Hurtz, G.M. & Donovan, J.J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869-879.

Jackson, D.N. (1994). *Jackson Personality Inventory. – Revised Manual*. Port Huron, MI: Sigma Assessment Systems.

Judge, T., Bono, J.E., Ilies, R. & Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.

Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-Factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*, 530-541.

McCrae, R.R. & Costa, P.T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.

McFarland, L.A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality tests scores. *International Journal of Selection and Assessment, 11*, 265-276.

Meade, A.W. (2004) Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*, 531-552.

Meehl, P.E. & Hathaway, S.R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30*, 525-564.

Mount, M. K. & Barrick, M. R. (1995). The Big Five personality dimensions: implications for research and practice in human resources management. In K.M. Rowland & G. Ferris (Eds.), *Research in personnel and human resources management, vol. 13* (pp 153-200). Greenwich, CT: JAI Press.

Mueller-Hanson, R., Heggestad, E.D. & Thornton, G.C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.

Ones, D.S. & Viswesvaran, C. (1998a). Integrity testing in organizations. In R.W. Griffin, A. O'Leary-Kelly & J. Collins (eds). *Dysfunctional behaviour in organizations: Vol. 2. Nonviolent behaviours in organizations (pp. 243-276)*. Greenwich, CT: JAI Press.

Ones, D.S. & Viswesvaran, C. (1998b). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269.

Ones, D.S. & Viswesvaran, C. (2001a). Personality at Work: criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts & R.T. Hogan (Eds), *Applied personality psychology* (pp. 63-92). Washington, DC: American Psychological Association.

Ones, D.S. & Viswesvaran, C. (2001b). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9*, 31-39.

Ones, D.S., Viswesvaran, C. & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.

Ones, D.S., Viswesvaran. C. & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validi-

ties: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology (Monograph), 78*, 679-703.

Ones, D.S., Viswesvaran. C. & Schmidt, F.L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17*, 19-38.

Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun & D.N. Jackson (Eds.). *Role of constructs in psychological and educational measurement* (pp 49-69). Mahwah, NJ: Lawrence Erlbaum Associates.

Rolland, J.P. & Mogenet, J.L. (2001). *Système de description en cinq dimensions (D5D). Manuel réservé aux psychologues*. Paris: Les Editions du Centre de Psychologie Appliquée.

Rosse, J.G., Stecher, M.D., Miller, J.L. & Levin, R.A. (1998). The impact response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.

Rothman, S., Meining, D. & Barrick, M.R. (2002). *Predicting performance using personality measures in South Africa*. Symposium on Personality at work in a cross-cultural context (D.S. Ones, Chair), Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.

Salgado, J.F. (1996) Análisis exploratorio y confirmatorio del inventario de personalidad de cinco factores (IP/5F). *Psicológica, 17*, 353-366.

Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*, 30-43.

Salgado, J.F. (1998). The Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance, 11*, 271-288.

Salgado, J.F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*, 117-125.

Salgado, J.F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76*, 323-346.

Salgado, J.F. (2004). *Moderator Effects of Job Complexity on the Big Five Validity*. Poster presented at the Conference of the Society for Industrial and Organizational Psychology, Chicago, Illinois.

Salgado, J.F. & De Fruyt, F. (2005). Personality in personnel selection. In A. Evers, O. Schmit-Voskuyl & N. Anderson (Eds). *Handbook of personnel selection*. Oxford, UK: Blackwell.

Salgado, J.F., Remeseiro, C. & Iglesias, M. (1996). Personality and Test-Taking Motivation. *Psicothema, 8*, 553-562.

Salgado, J.F. & Moscoso (2000). Autoeficacia y criterios organizacionales de desempeño. *Apuntes de Psicología, 18*, 179-191.

Seisdedos, N. (1988). Selección, cuestionarios y distorsión motivacional (actitud inteligente de adaptación). *Revista de Psicología del Trabajo y las Organizaciones, 4*, 3-33.

SHL (1999). *OPQ32 Manual and User's Guide*. Surrey: SHL.

Van Rooy, D. L. & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior, 65*, 71-95.

Viswesvaran, C., Ones, D.S. & Hough, L.M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment, 9*, 277-289.

Yoo, T.Y. & Ming, B.M. (2002). *A meta-analysis of the Big Five and performance in Korea*. Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada, April.

Papeles del Psicólogo, 2005. Vol. 26, pp. 129-137

ASSESSMENT OF SELF-REPORT RESPONSE DISTORSION BY MEANS OF THE MMPI-2

Héctor González Ordi and Iciar Iruarrizaga Díez
Madrid Complutense University

The Minnesota Multiphasic Personality Inventory 2, MMPI-2 is one of the best validated multi-scale measures for random responding, malingered psychopathology and defensiveness. The present paper presents in detail different possibilities, strategies and scales provided by the MMPI-2 for the detection of overreporting response style (F, Fb and F(p) scales, positive F-K index, and FBS and DsR scales), underreporting response style (L and K scales, negative F-K index, and S and Wsd scales), and random response style (cannot-say or "?", F, Fb, VRIN and TRIN scales).

El Inventario Multifásico de Personalidad de Minnesota 2, MMPI-2, es uno de los instrumentos multiescalares de amplio espectro mejor validados para explorar estilos de respuesta aleatorios, simulación de psicopatología y defensividad. En el presente artículo se exponen en detalle las distintas posibilidades, estrategias y escalas que proporciona el empleo del MMPI-2 para la evaluación de estilos de respuesta sobredimensionados (escalas F, Fb, F(p), índice F-K positivo, FBS y DsR), estilos de respuesta infradimensionados (escalas L, K, índice F-K negativo, S y Wsd) y estilos de respuesta aleatorios (escalas ?, F, Fb, VRIN y TRIN).

Lying, deception, and the concealment, distortion and twisting of information are behaviours essential to the human being (Martínez Selva, 2005), who pursues a variety of aims in the process of social interaction, such as causing a positive impression in others, benefiting oneself or others, avoiding a potential punishment or simply maintaining good social interaction by trying to avoid unnecessarily hurting others' feelings (Vrij, 2001).

Let us imagine for a moment that we had been invaded by aliens who had the power of omnipresence and were all-seeing and all-hearing, always told the truth about everything, and spent their time simply observing us and constantly interfering in our human conversations. We would undoubtedly be plunged into absolute chaos, powerless to deal with this type of "truth game" (indeed, this provides the basis for the plot of Fredric Brown's excellent science-fiction novel *Martians Go Home*).

The use of diverse strategies for distorting information in pursuit of a particular aim is a constituent part of the social interaction process, known, accepted and consented to by all parties, as long as those strategies are within the limits of what is socially admissible (Kashy & DePaulo, 1996).

The field of psychological assessment through self-report is by no means immune to the distortion of information by respondents for various reasons, which is commonly referred to as response distortion (Miguel-Tobal, 1993; Baer, Rinaldo & Berry, 2003). The study of response distortion and the most effective strategies or instruments for detecting it is strongly on the increase, and as it develops it is having more and more important consequences for clinical, forensic and medico-legal practice.

Among the different types of response distortion found are the following (Baer, Rinaldo & Berry, 2003):

1. "Bad image" patterns, overreporting response styles or malingering (*faking bad*), when the respondent deliberately tries to create the impression of having some disorder or deterioration through the exaggeration or fabrication of symptoms and problems and by emphasizing as far as possible his or her negative characteristics.
2. "Good image" patterns, underreporting response styles, defensiveness or social desirability (*faking good*), when respondents deliberately attempt to create a favourable impression of themselves, omitting to mention, denying or concealing symptoms

Correspondence: Dr. Héctor González Ordi. *Departamento de Psicología Básica II (Procesos Cognitivos). Facultad de Psicología. Universidad Complutense de Madrid. Campus de Somosaguas s/n. 28223 Madrid. Spain. E-mail: hectorgo@psi.ucm.es*

and problems, and highlighting their positive characteristics.

3. Random response style, when the subject responds independently of the item content, due to difficulties in reading or understanding items, reluctance to cooperate, carelessness, lack of concentration or confused states of mind. Within this category are the "acquiescence" and "non-acquiescence" approaches, which involve the tendency to responder indiscriminately "true" or "false" to all the items, regardless of their content.

THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY: MMPI, MMPI-2 AND MMPI-A

The Minnesota Multiphasic Personality Inventory (MMPI), originally developed by Hathaway and McKinley (1940), and its subsequent revised and restandardized versions for adults, the **MMPI-2** (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989) and for adolescents, the **MMPI-A** (Butcher, Williams, Graham, Archer, Tellegen, Ben-Porath & Kaemmer, 1992), published in the late 1980s, is one of the most widely used questionnaires for assessing psychopathological disorders in the clinical field in general (Lubin, Larsen & Matarazzo, 1984; Piotrowski, 1998) and in the forensic context in particular (Bartol & Bartol, 2004; Boccaccini & Brodsky, 1999).

As we shall see, the MMPI-2 includes various indicators of its validity that have demonstrated their utility in the detection of faking (Elhai, Naifeh, Zucker, Gold, Deitsch & Frueh, 2004; Guriel & Fremouw, 2003; Rogers, Sewell, Martin & Vitacco, 2003). Indeed, according to Rogers (1997), the MMPI and MMPI-2 are the most well-validated wide-ranging multi-scales instruments for exploring random response styles, psychopathological malingering and defensiveness.

Butcher and Ben-Porath (2004) list some of the characteristics that contribute to the popularity and extensive use of this wide-ranging psychopathological assessment instrument over its more than sixty years of existence: (1) it includes a large quantity of psychopathological and personality factors that have shown themselves to be reliable, valid and stable over time; (2) it has incorporated new scales to take account of conceptual advances in psychopathology, thus becoming periodically renewed and updated; (3) it permits individual profiles to be checked against an extensive database built up over decades of research; (4) it permits objective interpretation, following standardized norms; and (5) it has been translated and adapted for several languages and countries, thus making possible cross-cultural comparison.

The Spanish adaptation of the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1999) includes 7 validity scales, the 10 original basic clinical scales of the MMPI with their 31 specific subscales, 15 content scales and 15 supplementary scales by various authors, which have been added to the instrument over the years. In total, 78 scales and subscales, making the MMPI truly unique in terms of the richness, scope and diversity of the information it provides, as can be seen in Table 1.

PROCEDURE FOR ASSESSMENT OF RESPONSE DISTORSION BY MEANS OF THE MMPI-2

The protocol for assessment of response distortion we shall follow in this article is based on the steps for assessment of MMPI-2 validity proposed by Greene (1997), which essentially consist of five phases, as can be seen in Table 2: once the MMPI-2 has been administered and filled out, the number of omissions (unanswered items) and mistaken responses made by the subject are detected. After this, a rating is given to the consistency and reliability of the responses, and as long as the distortions found do not advise to the contrary, the assessor proceeds to the clinical interpretation of the basic scales and their subscales, content scales and supplementary scales. We shall now consider each of these five phases in more detail.

Administration of the MMPI-2

The MMPI-2, in its complete or standard version (the most widely used and recommended), is a 567-item questionnaire, with a dichotomic true-false response format, designed for application to adults (?18 years), and with an estimated administration time of between 1 and 2 hours for the majority of cases. In patients with severe psychopathology this administration time may extend to between 3 and 4 hours. Exceptionally, there is an abbreviated form of application in which only items 1 to 370 are administered, though its use is not normally advisable, since it only permits the assessor to obtain reliable results for the basic clinical scales and the validity scales (Nichols, 2001).

Detection of omissions or mistaken responses

Once the MMPI-2 has been filled out by the respondent, the first step in the assessment of response distortion is to detect the number of omissions or mistaken responses the

TABLE 1: MMPI-2
SCALES AND SUBSCALES IN THE SPANISH ADAPTATION (MODIFIED FROM GONZALEZ ORDI & GOMEZ SEGURA, 2002)

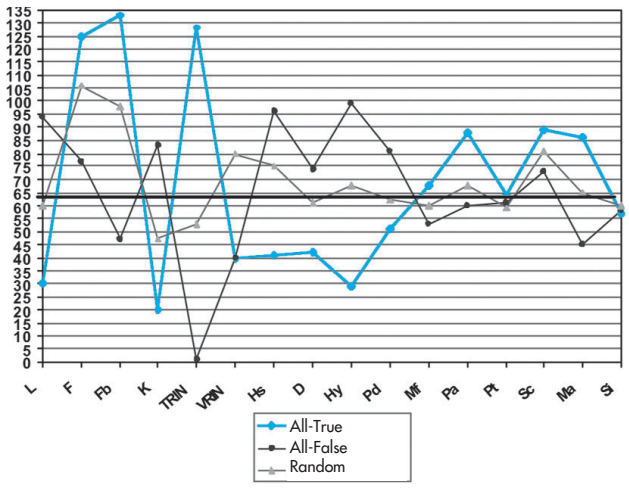
VALIDITY SCALES ?= Cannot-say K= Correction (Subtle defensiveness) TRIN= True Response Inconsistency L= Lie F(b)= Back Infrequency F= Infrequency (Exaggeration of symptoms) VRIN= Variable Response Inconsistency	
BASIC CLINICAL SCALES	**Harris & Lingoes Subscales**
1. Hs= Hypochondria	
2. D= Depression	D1= Subjective depression D2= Psychomotor retardation D3= Physical malfunctioning D4= Mental dullness D5= Brooding
3. Hy= Hysteria	Hy1= Denial of social anxiety Hy2= Need for affection Hy3= Lassitude-malaise Hy4= Somatic complaints Hy5= Inhibition of aggression
4. Pd= Psychopathic deviate	Pd1= Familial discord Pd2= Authority Problems Pd3= Social imperturbability Pd4= Social alienation Pd5= Self-alienation
5. Mf= Masculinity-Femininity	
6. Pa= Paranoia	Pa1= Persecutory ideas Pa2= Poignancy Pa3= Naivete
7. Pt= Psychasthenia	
8. Sc= Schizophrenia	Sc1= Social alienation Sc2= Emotional alienation Sc3= Lack of ego mastery, cognitive Sc4= Lack of ego mastery, conative Sc5= Lack of ego mastery, defective inhibition Sc6= Bizarre sensory experiences
9. Ma= Hypomania	Ma1= Amorality Ma2= Psychomotor acceleration Ma3= Imperturbability Ma4= Ego inflation
0. Si= Social Introversion	**"Si" Subscales** Si1= Shyness/self-consciousness Si2= Social avoidance Si3= Alienation—self and others
CONTENT SCALES ANX= Anxiety OBS= Obsessiveness HEA= Health concerns ANG= Anger ASP= Antisocial practices LSE= Low self-esteem FAM= Family problems TRT= Negative treatment indicators FRS= Fears DEP= Depression BIZ= Bizarre mentation CYN= Cynicism PA= Type A Behaviour SOD= Social discomfort WRK= Work interference	
SUPPLEMENTARY SCALES A= Anxiety MAC-R= MacAndrew Alcoholism Scale-Revised Re= Social responsibility GF= Feminine gender role MDS= Marital distress scale R= Repression O-H= Overcontrolled hostility Mt= College maladjustment PK= Post-traumatic stress disorder scale APS= Addiction potential scale Es= Ego strength Do= Dominance GM= Masculine gender role PS= Post-traumatic stress disorder scale AAS= Addiction acknowledgement scale	

respondent has made, by means of the "?" Scale. Given the length of the instrument, it is frequent for the majority of subjects, whether they present psychopathology or not, to fail to respond to some items or to erroneously mark both responses, true and false. Indeed, Greene (1997) has estimated the expectable range of omissions at between 1 and 15 for normal subjects and between 0-20 for psychopathological patients. In general, the administration protocol is considered to be invalid if the respondent leaves 30 or more items unanswered in the first 370; if these omissions occur after item 370, clinical interpretation can go ahead for the basic clinical scales and validity scales, but not for the rest of the scales. Excessive omission of items is usually considered to be related to patterns of defensiveness, indecision, carelessness, fatigue or inability to read and understand the items (Butcher & Williams, 1992; Graham, 1993).

TABLE 2
PROCEDURE FOR ASSESSMENT OF RESPONSE DISTORTION BY MEANS OF THE MMPI-2 (BASED ON GREENE, 1997)

PHASES	OBJECTIVES	SCALES
Phase 1	Administration of the MMPI-2	Standard: 567 items Abbreviated: 370 items
Phase 2	Detection of omissions and mistaken responses	Cannot-say scale (?)
Phase 3	Assessment of consistency of responses	Random profiles, VRIN and TRIN Scales, F and Fb Scales, tendency to reply true or false
Phase 4	Assessment of reliability of responses 1. Overreporting response patterns 2. Underreporting response patterns	 F, Fb Scales, F-K Index L, K Scales, F-K Index
Phase 5	Clinical interpretation of the MMPI-2	Basic clinical scales and their subscales, content scales and supplementary scales

FIGURE 1
RESPONSE DISTORTION PROFILES IN THE MMPI-2: "ALL-TRUE", "ALL-FALSE" AND "RANDOM" (IN T SCORES)



Assessment of consistency of responses

Once it has been confirmed that the number of omissions and mistaken responses is within the acceptable limits for ensuring the validity of the protocol, the next phase in the assessment of response distortion in the MMPI-2 involves examining whether the subject has responded consistently to the items. Subjects can respond inconsistently to items in various ways: tending to answer "true" (acquiescence), tending to answer "false" (non-acquiescence), or simply responding randomly. The distortion profiles obtained in the MMPI-2 as a result of these three forms of inconsistent response can be seen in Figure 1.

One of the most sensitive scales of the MMPI-2 for detecting patterns of inconsistent response is the Infrequency or F Scale (Clark, Gironda & Young, 2003; Sewell & Rogers, 1994), which, as can be seen in Figure 1, appears as unusually high (above the normative cut-off point, T=65) for the three forms of inconsistent response. The F Scale, and its partner the Back Infrequency or Fb Scale, are instruments designed to detect infrequent response, or "true" responses to items that would receive a "true" response from less than 10% of the normative population; thus, high scores on the F and Fb Scales (T?65) would indicate a significant deviation from normative patterns and a preponderance of non-conventional response styles (Nichols, 2001).

Having confirmed a significantly high score on the F Scale, it remains to identify the direction of the inconsistent response pattern. The TRIN and VRIN are extremely useful for discriminating the characteristics of the supposed inconsistent response pattern.

The TRIN Scale (*True Response Inconsistency Scale*) is designed for detecting whether there is an acquiescent (tendency to reply "true") or non-acquiescent (tendency

to reply "false") distorted pattern of responses. As can be seen in Figure 1, the "all-true" response profile is characterized by markedly high scores on the TRIN Scale, while for the "all-false" response profile TRIN scores are extremely low. The VRIN Scale (*Variable Response Inconsistency Scale*), on the other hand, is designed to specifically detect random response styles, inconsistent with the item content. Indeed, Figure 1 shows clearly how the VRIN score is only unusually high in the case of the "random" response profile, and not in the cases of "all-true" or "all-false". Extreme TRIN scores confirm that the subject has responded to the instrument in a "careless" way, without telling us precisely whether his/her response to the content of the items was consistent or not.

Assessment of reliability of responses

The aim of assessing the reliability of responses is to identify the presence or absence of distorted patterns of response that hinder the correct clinical interpretation of the MMPI-2. Basically, there are two types of pattern to consider in this regard: (1) "Bad image" patterns, overreporting response styles or malingering (*faking bad*), and (2) "Good image" patterns, underreporting response styles, defensiveness or social desirability (*faking good*).

Overreporting response styles

The Infrequency Scales F and Fb (Back Infrequency) have demonstrated their utility for effectively identifying individuals who attempt to present themselves in a bad light, deliberately malingering psychopathological symptoms (Bury & Bagby, 2002; Elhai, Naifeh, Zucker, Gold, Deitsch & Frueh, 2004; Graham, Watts & Timbrook, 1991; Nicholson, Mouton, Bagby, Buis, Peterson & Buidas, 1997; Strong, Greene & Schinka, 2000). In fact, these scales contain items selected for detecting atypical or unusual response styles, whose content refers to bizarre or unusual symptoms of severe psychopathology (Nichols, 2001). As Greene (1997) rightly points out, high scores on these scales may be due to the presence of inconsistent response styles (as we saw in the previous section), to the existence of actual severe psychopathology, or to a pattern of simulation of responses, in which case scores on the basic clinical MMPI-2 scales would be inflated. Low scores would tend to be associated with absence of genuine psychopathology, or with patterns of defensiveness, deflating the scores on the basic clinical MMPI-2 scales. With regard to "faking bad" or overreporting patterns, Butcher (2005) recommends considering the presence of malingering of symptoms when F and/or Fb present T scores of over 100, and VRIN is less than or equal to 79.

Another relevant indicator of faking is Gough's F-K index (1950). This index is obtained by subtracting the raw score on the K validity Scale from the raw score on the F validity Scale (F minus K). If the index is positive after a given cut-off point, the subject will display a tendency to *fake bad*, or deliberately exaggerate symptoms; if the index is negative after a given cut-off point, the subject will show a tendency to the denial or concealment of symptoms – defensiveness or *faking good*.

One of the problems with the F-K index is that there is no consensus among authors in relation to the definitive cut-off points recommended for effectively distinguishing malingerers from non-malingerers, since these cut-off points depend to a large extent on the measures used for obtaining them. Indeed, the scientific literature refers to cut-off points recommended for malingering ranging from +6 to +27, and for defensiveness of between -11 and -20, all in North American samples (see Butcher & Williams, 1992; Greene, 1997; Meyers, Millis & Volkert, 2002; Nichols, 2001; Pope Butcher & Seelen, 1993). As regards the use of the F-K in our own country, Spain, recommended cut-off points have been calculated specifically for malingering and defensiveness for both the MMPI-2 (González Ordi & Gómez Segura, 2002) and the MMPI-A (González Ordi, 2005), based on the samples of reference used for the Spanish adaptation of the two instruments.

Despite the fact that recent research suggests it is no more effective in the detection of faking than the F alone (Bury & Bagby, 2002; Butcher, 2005; Nicholson et al., 1997), the F-K index is sufficiently sensitive to the detection of malingering (it in fact functions much better in this task than in the assessment of defensiveness or denial of symptoms, according to Nichols, 2001) to be worth continuing to take into account as providing additional information, especially as it correlates positively and significantly with the latest generation of self-report instruments for the assessment of malingering, such as the SIMS - *Structured Inventory of Malingered Symptomatology* (Widows & Smith, 2005), and is still widely used in the field of forensic assessment as an aid to detecting the deliberate exaggeration of psychopathological symptoms (Ben-Porath, Graham, Hall, Hirschman & Zaragoza, 1995; González Ordi & Gancedo Rojí, 1999).

Underreporting response styles

The MMPI-2 indices most widely used for assessing underreporting response styles are the L and K Scales (Baer & Miller, 2002).

The *Lie* or L scale consists of 15 items selected with the aim of identifying respondents who deliberately try to present a defensive pattern of responses, in the sense of concealing the most negative aspects of their personality, especially if they obtain T scores of over 66 (Butcher, 2005). T scores of between 60 and 65 would reflect an attempt by the subject to present as favourable an image of him or herself as possible (hiding problems of personal adjustment or the truth), an inability to admit mild moral transgressions and an excessive sense of virtue and morality (Butcher & Williams, 1992; Graham, 1993).

The K scale was developed as a measure of defensiveness and as a factor for correcting the tendency of subjects to deny the presence of psychopathological problems (Butcher, 2005). As a correcting factor it is applied at different values to the basic clinical scales Hs, Pd, Pt, Sc and Ma, for adjusting their final score. As a scale of independent validity, when K presents T scores between 60 and 69 it reflects subjects' tendency to display a favourable image of themselves, minimizing their problems as far as possible; when K presents T scores of 70 or over, it is reasonable to consider that the subject presents a defensive response pattern (Butcher & Williams, 1992; Pope. Butcher & Seelen, 1993).

Finally, the F-K index can also be useful as additional information on subjects' tendency to underreport in their responses to the MMPI-2, as mentioned above.

It is important to point out here that while the scales designed to explore the tendency to overreport in responding to the MMPI-2 (F, Fb, positive F-K index) have received greater research interest, and enjoy more substantial empirical support for their effectiveness in detecting the deliberate exaggeration of psychopathological symptoms and correctly classifying malingerers (*bad fakers*) from non-malingerers, the scales designed for detecting the tendency to present oneself in an exaggeratedly favourable light, dissimulating or concealing symptoms or psychopathological problems (L, K, negative F-K index), do not have such unanimous and generalized empirical support, so that more research effort is required (see Baer & Miller, 2002).

SCALES DERIVED FROM THE MMPI-2 FOR THE ASSESSMENT OF RESPONSE DISTORTION

In addition to the validity scales routinely included in the Spanish version of the MMPI-2, there are a number of scales derived empirically from the MMPI-2 itself, but which did not originally form part of it, and are currently used as sources of additional information for the assessment of response distortion patterns.

Infrequency-Psychopathological Scale [F(p)]

This F(p) Scale (Arbisi & Ben-Porath, 1995) was created as an additional measure of validity for explaining more specifically the high scores found on the MMPI-2 F validity scale. In fact, Arbisi and Ben-Porath (1995) suggest that when F and F(p) present high scores, it is more reasonable to attribute such high scores to a pattern of response simulation than to the presence of actual severe psychopathology, especially if the VRIN and TRIN scale scores are not significantly high. Thus, considering the F and F(p) scales jointly will be more effective for distinguishing between groups with genuine psychopathology and groups of malingerers than using the F scale alone (Bury & Bagby, 2002; Rothke, Friedman, Jaffe, Greene, Wetter, Cole & Baker, 2000; Storm & Graham, 2000; Strong, Greene & Schinka, 2000).

Fake Bad Scale (FBS)

The FBS Scale (Less-Haley, English & Glenn, 1991) was designed specifically with the aim of helping to detect malingering of somatic complaints in the forensic context. It includes items referring to somatic symptoms, sleep disorders, symptoms related to tension and stress, lack of energy, anhedonia, and so on. Although there was a fair amount of research on this scale as a possible instrument for the detection of malingering during the 1990s, recent studies advise against its use as a scale for detecting patterns of malingering, arguing that it is more appropriate as a scale that assesses the tendency for expressing severe psychopathological symptomatology, focusing on more somatic aspects and emotional distress (Butcher, Arbisi, Atlis & McNulty, 2003).

Revised Gough Dissimulation Scale [DsR]

The Revised Gough Dissimulation Scale (Gough, 1957) (DsR) has been employed in the forensic field for distinguishing between subjects who malinger neurotic symptoms, patients with genuine symptoms and normal population. Although less widely employed than other

scales for detecting overreporting response styles, such as the F(p), it is still used as an additional indicator of possible malingering (Bury & Begby, 2002; Storm & Graham, 2000).

Superlative Self-Presentation Scale (S)

The Superlative Self-Presentation or S Scale (Butcher & Han, 1995) was designed to detect subjects who present themselves in a superlative way, exaggeratedly highlighting their positive aspects. In fact, it correlates positively and significantly with the K validity scale (Greene, 1997), providing additional information on the tendency to present a favourable image of oneself, dissimulating or concealing psychopathological symptoms or problems. Thus, Butcher (2005) suggests that when the S scale S presents typical scores over 70 assessors should consider the possible presence of a defensive response pattern.

Social Desirability Scale (Wsd)

The Social desirability or Wsd Scale (Wiggins, 1959) is a classic instrument in the history of the MMPI, and was designed to assess the tendency to present oneself in a socially desirable way. It is one of the scales most traditionally used for exploring underreporting response styles or patterns of defensiveness.

We have tried in this article to explore the possibilities offered by the MMPI-2 for detecting and assessing response distortion and faking. Throughout the last 65 years, the MMPI and its re-standardization, the MMPI-2, as instruments for the assessment of psychopathology, have been constantly changing and renewing themselves, and have paid particular attention to the design of self-report-based strategies for detecting response distortion, which have had, and continue to have, important applied implications for the clinical, forensic and medico-legal contexts.

The MMPI-2 currently offers multiple possibilities for the assessment of overreporting response styles (F, Fb, F(p) scales, positive F-K index, FBS and DsR), underreporting response styles (L, K scales, negative F-K index, S and Wsd scales) and random response styles ("?", F, Fb, VRIN and TRIN scales). Use of the information deriving from these scales facilitates detection of the response distortion that may occur when a subject is administered this instrument, especially if that subject has the intention of faking. However, to definitively establish the presence of response distortion, the professional should take into account other information sources, as well as the MMPI-2, since the study of faking necessarily requires detailed multimethod/multisystem psychological assessment (González Ordi & Gancedo Rojí, 1999).

REFERENCES

Arbisi, P.A. & Ben-Porath, Y.S. (1995). On MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequent Psychopathology Scale F(p). *Psychological Assessment, 7*, 424-431.

Baer, R.A. & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: a meta-analytic review. *Psychological Assessment, 14,* 16-26.

Baer, R.A., Rinaldo, J.C. & Berry, D.T.R. (2003). Self-report distortions (including faking, lying, malingering, social desirability). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment. Vol. 2* (pp. 861-866). London: SAGE.

Bartol, C.R. & Bartol, A.M. (2004). *Introduction to forensic psychology*. London: SAGE. Ben-Porath, Y.S., Graham, J.R., Hall, G., Hirschman, R. & Zaragoza, M.S. (Eds.) (1995). *Forensic applications of the MMPI-2.* London: SAGE.

Boccaccini, M.T. & Brodsky, S. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice, 30*, 253-259.

Bury, A.S. & Bagby, R.M. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment, 14*, 472-484.

Butcher, J.N. (2005). Assessing MMPI-2 profile validity. In G.P. Koocher, J.C. Norcross & S.S. Hill III (Eds.), *Psychologists' desk reference*. 2nd edition (pp. 128-132). New York: Oxford University Press.

Butcher, J.N., Arbisi, P.A., Atlis, M.M. & McNulty, J.L. (2003). The construct validity of the Less-Haley Fake Bad Scale. Does this scale measuring somatic malingering or feigned emotional distress? *Archives of Clinical Neuropsychology, 18,* 473-485.

Butcher, J.N. & Ben-Porath, Y.S. (2004). Use of the MMPI-2 in medico-legal evaluations: An alternative interpretation for the Senior and Douglas (2001) critique. *Australian Psychologist, 39*, 44-50.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. & Kaemmer, B. (1989). *MMPI-2: Manual of administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. & Kaemmer, B. (1999). *MMPI-2: Inventario Multifásico de Personalidad de Minnesota-2. Manual.* (Spanish adaptación by A. Avila-Espada & F. Jiménez-Gómez). Madrid: TEA Ediciones.

Butcher, J.N. & Han, K. (1995). Development of a MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. In J.N. Butcher & C.D. Spielberger (Eds.), *Advances in personality assessment. Vol. 10* (pp. 25-50). Hillsdale, NJ: Erlbaum.

Butcher, J.N. & Williams, C.L. (1992). *Essentials of MMPI-2 and MMPI-A interpretation.* Minneapolis, MN: University of Minnesota Press.

Butcher, J.N., Williams, C.L., Graham, J.R., Archer, R.P., Tellegen, A., Ben-Porath, Y.S. & Kaemmer, B. (1992). *MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Clark, M.E., Gironda, R.J. & Young, R.W. (2003). Detection of back random responding: effectiveness of MMPI-2 and Personality Assessment Inventory validity indices. *Psychological Assessment, 15,* 223-234.

Elhai, J.D., Naifeh, J.A., Zucker, I.S., Gold, S.V., Deitsch, S.E. & Frueh, B.C. (2004). Discriminating malingered from genuine civilian posttraumatic stress disorder: A validation of the three MMPI-2 infrequency scales (F, Fp, and Fptsd). *Assessment, 11,* 139-144.

González Ordi, H. (2005). *Informe interpretativo informatizado del Inventario Multifásico de Personalidad de Minnesota para adolescentes (MMPI-A): Manual.* Madrid: TEA Ediciones.

González Ordi, H. & Gómez Segura, J. (2002). *Informe interpretativo informatizado del Inventario Multifásico de Personalidad de Minnesota-2 (MMPI-2): Manual.* Madrid: TEA Ediciones.

González Ordi, H. & Gancedo Rojí, M. (1999). Evaluación de la simulación de respuesta: diagnóstico, técnicas y procedimientos. In A. Ruano Hernández, J.M. Muñoz Céspedes & C. Cid Rojo (Coords.), *Psicología de la rehabilitación* (pp. 405-441). Madrid: Fundación MAPFRE-Medicina.

Gough, H.G. (1950). The F minus K dissimulation index for the MMPI. *Journal of Consulting Psychology, 14,* 408-413.

Gough, H.G. (1957). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Graham, J.R. (1993). *MMPI-2: Assessing personality and psychopathology.* New York: Oxford University Press.

Graham, J.R., Watts, D. & Timbrook, R.E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment, 57,* 264-277.

Greene, R.L. (1997). Assessment of malingering and defensiveness by multiscale personality inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-207). New York: The Guilford Press.

Guriel, J. & Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A clinical review. *Clinical Psychology Review, 23,* 881-904.

Hathaway, S.R. & McKinley, J.C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10,* 249-254.

Kashy, D.A. & DePaulo, B.M. (1996). Who lies? *Journal of Personality and Social Psychology, 70,* 1037-1051.

Less-Haley, P.R., English, L.T. & Glenn, W.J. (1991). A fake bad scale on the MMPI-2 for personal-injury claimants. *Psychological Reports, 68,* 203-210.

Lubin, B., Larsen, R.M. & Matarazzo, J.D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist, 39,* 451-454.

Martínez Selva, J.M. (2005). *La psicología de la mentira.* Barcelona: Paidós.

Meyers, J.E., Millis, S.R. & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology, 17,* 157-169.

Miguel-Tobal, J.J. (1993). Cuestionarios, inventarios y escalas. In F.J. Labrador, J.A. Cruzado & M. Muñoz (Eds.), *Manual de técnicas de modificación y terapia de conducta* (pp. 151-180). Madrid: Pirámide.

Nichols, D.S. (2001). *Essentials of MMPI-2 assessment.* New York: John Wiley & Sons.

Nicholson, R.A., Mouton, G.J., Bagby, R.M., Buis, T., Peterson, S.A. & Buigas, R.A. (1997). Utility of MMPI-2 indicators of response distortion: receiver operating characteristics analysis. *Psychological Assessment, 9,* 471-479.

Piotrowski, C. (1998). Assessment of pain: A survey of practicing clinicians. *Perceptual and Motor Skills, 86,* 181-182.

Pope, K.S., Butcher, J.N. & Seelen, J. (1993). *The MMPI, MMPI-2 & MMPI-A in court. A practical guide for expert witnesses and attorneys.* Washington, DC: American Psychological Association.

Rogers, R. (1997). Current status of clinical methods. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 373-397). New York: The Guilford Press.

Rogers, R., Sewell, K.W., Martin, M.A. & Vitacco, M.J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10,* 160-177.

Rothke, S.E., Friedman, A.F., Jaffe, A.M., Greene, R.L., Wetter, M.W., Cole, P. & Baker, K. (2000). Normative data for the F(p) scale of the MMPI-2: implications for clinical and forensic assessment of malingering. *Psychological Assessment, 12,* 335-340.

Sewell, K.W. & Rogers, R. (1994). Response consistency and the MMPI-2: Development of a simplified screening scale. *Assessment, 1,* 293-299.

Strong, D.R., Greene, R.L. & Schinka, J.A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings. *Psychological Assessment, 12,* 166-173.

Storm, J. & Graham, J.R. (2000). Detection of coached general malingering on the MMPI-2. *Psychological Assessment, 12,* 158-165.

Vrij, A. (2001). *Detecting lies and deceit.* New York: John Wiley & Sons.

Widows, M.R. & Smith, G.P. (2005). *SIMS: Structured Inventory of Malingered Symptomatology. Professional manual.* Lutz, FL: PAR: Psychological Assessment Resources.

Wiggins, J.S. (1959). Interrelationship among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23,* 419-427.