INTERNATIONAL JOURNAL OF PSYCHOLOGY, 2006, 41 (4), 304–319

Psychology Press
Taylor & Francis Group

Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence

Jaume Masip, Eugenio Garrido, and Carmen Herrero

University of Salamanca, Spain

Research into the nonverbal detection of deception has typically been conducted by asking observers to judge whether a number of videotaped statements are truthful or deceptive. In most cases, the behavioural segments used in this research have been very short. A typical result is that observers tend to judge the statements as truthful (truth bias). In the present experiment, observers watched a series of video clips showing senders answering three questions about an event that they had witnessed. Observers had to indicate whether each sender's statement was truthful or deceptive, their judgmental confidence, and when they had made their decision about the sender's credibility: during his or her first, second, or third answer. Competing predictions were made about the influence of the decision moment on the observers' judgments and accuracy. The results replicated most research findings reported in the US and North-European literature, including the truth bias phenomenon. However, the proportion of judgments of truthfulness decreased as observers decided later, particularly for the deceptive statements. This yielded an increase in accuracy in judging deceptive accounts. These results are consistent with the idea that initial credibility judgments are made heuristically, either because there is not enough information available or because observers are in the first, automatic stage of current attribution and person perception models. Heuristic decision making may produce a high proportion of judgments of truthfulness. Later judgments would be made in a systematic manner. The truth bias detected in deception research may be caused by researchers having used very brief and uninformative behavioural samples. The moment when observers made their decision had only a marginal negative influence on confidence.

La recherche sur la détection non verbale de la duperie a typiquement été conduite en demandant à des observateurs de juger dans quelle mesure un certain nombre d'affirmations filmées sont véridiques ou mensongères. Dans la plupart des cas, les segments comportementaux utilisés ont été très courts. Généralement, les résultats ont montré que les observateurs ont tendance à juger les affirmations comme étant véridiques (biais de vérité). Dans la présente expérimentation, les observateurs ont visionné une série de vidéos montrant des individus répondant à trois questions concernant un événement dont ils ont été témoins. Les observateurs devaient indiquer dans quelle mesure chaque affirmation des individus étaient véridique ou mensongère, leur niveau de confiance en leur jugement et à quel moment ils ont pris leur décision à propos de la crédibilité de l'individu: durant la première, la deuxième ou la troisième réponse. Des prédictions concurrentes ont été avancées concernant l'influence du moment de décision sur les jugements et l'exactitude des observateurs. Les résultats obtenus ont confirmé la plupart des résultats de recherche rapportés dans les écrits états-uniens et nord européens, incluant le phénomène du biais de vérité. Cependant, la proportion de jugements de véracité diminuait alors que le temps de prise de décision augmentait, ceci étant particulièrement le cas pour les affirmations mensongères. Ceci a mené à une augmentation de l'exactitude dans le jugement des mensonges. Ces résultats sont en accord avec l'idée que les jugements initiaux de crédibilité sont le produit d'heuristiques, soit parce qu'il n'y a pas suffisamment d'informations disponibles ou parce que les observateurs se trouvent dans un premier stage automatique d'attribution du moment et de modèles de perception de la personne. La prise de décision

Correspondence should be addressed to Jaume Masip, Department of Social Psychology and Anthropology, University of Salamanca, Facultad de Psicología, Avda. de la Merced, 109–131, 37005 Salamanca, Spain (E-mail: jmasip@usal.es).

The authors are grateful to the research assistants Gema Martín and Monica Sánchez, to Yago Reis and the actors Manolo Muñoz, Marta Calzada and José Romo, to the technician Javier Tamames, to José L. Vega, the Dean of the Faculty of Psychology of the University of Salamanca, for permitting us to use his office to record the tapes, and to two anonymous reviewers for their helpful comments on an earlier draft of the manuscript. This research was supported by the Junta de Castilla y León, Programa de Apoyo a Proyectos de Investigación, Ref. SA023/03.

http://www.psypress.com/ijp DOI: 10.1080/00207590500343612

heuristique peut produire une proportion élevée de jugements de véracité. Les derniers jugements peuvent être faits de manière systématique. Le biais de vérité détecté dans la recherche sur la duperie peut être causé par les chercheurs qui ont utilisé de très brefs échantillons comportementaux contenant des informations limitées. Le moment auquel les observateurs prennent leur décision a eu une influence seulement marginalement négative sur la confiance des observateurs.

Normalmente, la investigación sobre la detección no-verbal del engaño se ha realizado pidiendo a unos observadores que juzguen si una serie de declaraciones en vídeo son verdaderas o falsas. En la mayoría de los casos, los segmentos conductuales empleados en esa investigación han sido muy cortos. Un resultado habitual es que los observadores tienden a juzgar las declaraciones como verdaderas (sesgo de veracidad). En el presente experimento, los observadores visualizaron una serie de fragmentos de vídeo en que aparecían unos emisores respondiendo a tres preguntas sobre un acontecimiento que habían presenciado. Los observadores tuvieron que indicar si la declaración de cada emisor era verdadera o falsa, su confianza en este juicio, y cuándo habían hecho su decisión sobre la credibilidad: durante su primera respuesta, la segunda, o la tercera. Se hicieron predicciones mutuamente excluyentes sobre la influencia del momento de decisión sobre los juicios y la precisión de los observadores. Los resultados replicaron la mayoría de los hallazgos de la investigación estadounidense y noreuropea, incluyendo el fenómeno del sesgo de veracidad. Sin embargo, la proporción de juicios de verdad decreció a medida que los observadores respondían más tarde, especialmente para las declaraciones falsas. Esto generó un aumento en la precisión obtenida al juzgar tales declaraciones. Estos resultados son consistentes con la idea de que los juicios de credibilidad iniciales se hacen heurísticamente, ya sea porque la información accesible es insuficiente o porque los observadores están en el primer paso, automático, de los modelos actuales de atribución y percepción de personas. La toma de decisiones heurística puede producir una alta proporción de juicios de verdad. Los juicios posteriores se harían de forma sistemática. El sesgo de veracidad detectado en la investigación del engaño puede deberse al empleo de muestras conductuales muy breves y poco informativas. El momento en que los observadores hicieron su decisión sólo tuvo una influencia marginal de signo negativo sobre la confianza.

The nonverbal detection of deceit has drawn the attention of many researchers over the last decades. In the typical deception-detection experiment, observers have to watch or listen to either recorded or live audio, visual, or audiovisual communications of a number of witnesses or suspects (the communication senders). After each communication, they have to indicate, normally in a form, whether the sender was lying or telling the truth (e.g., Miller & Stiff, 1993).

Using this procedure, researchers have studied the influence of a number of variables on observers' credibility judgments and their accuracy in detecting deceit (see Vrij, 2000, for an overview). However, a variable that, very surprisingly, has received little attention is the moment at which observers make their decision about the senders' veracity. Indeed, as time goes by, the communication sender provides progressively more information through his or her verbal and nonverbal behaviour.[1] Therefore, the later observers make their judgments, the more information they will have at their disposal on which to base those judgments. This might have an impact on observers' judgments and accuracy. However, very often deception scholars have conducted their experiments using as stimuli videotapes that lasted only between 10 or 20 s (Feldman, Jenkins, & Popoola, 1979; Zuckerman, DeFrank, Hall, Larrace, & Rosenthal, 1979) and 60 s (Frank & Ekman, 1997). Although there are some cases in which the behavioural segments have been longer, only rarely have they lasted more than 120 s. Thus, not only have researchers neglected the influence of the moment at which observers make their decisions on their credibility judgments, but they have also used videotapes so brief that one may wonder whether they can contain enough information for observers to make a reasoned judgment. This might account for the finding that observers' overall accuracy in detecting truthful and deceptive performances is only slightly, and often nonsignificantly, above chance. Typically, accuracy falls within the 45–60% correct judgments, where 50% is chance accuracy (for reviews on observers' detection accuracy, see

[1]It could be argued that if the sender chooses not to communicate, then more time does not result in an increase in information. However, as Watzlawick, Beavin, and Jackson (1967) stated, it is not possible not to communicate, because even silence and stillness are informative. A sender trying not to convey information might, for instance, take a long time before answering questions, speak slowly, pause often, provide few descriptive details, and inhibit his or her body movements. Latency period, speech rate, number and length of pauses, quantity of details, and inhibitions of movements might be actual, perceived, or stereotypical deception cues (e.g., Vrij, 2000). Therefore, they are often very informative.

Bond & DePaulo, in press; DePaulo, Zuckerman, & Rosenthal, 1980; Kalbfleisch, 1985; Vrij, 2000). It is surprising that, among the many explanations that have been provided to account for this limited accuracy (see Kraut, 1980; O'Sullivan, 2003; Vrij, 2000), as far as the present authors know no-one has ever suggested that judgments may only be random, owing to the scarce information that can be contained in the brief behavioural samples used in deception research.

In the present report, a study is described that was designed to examine the influence of the observers' decision-making moment on their judgments, accuracy, and confidence. A number of senders were shown a videotape. Later on, they were interviewed about the facts depicted in the videotape. They had to answer three questions twice, once truthfully and once deceptively, while being videotaped. Each sender's answer to the first question was considered as Moment 1, the answer to the second question as Moment 2, and the answer to the third question as Moment 3. The observers watched each sender's performance on videotape. They had to write whether they thought the sender was lying or telling the truth, their confidence in that judgment, and when they had come to the conclusion about the sender's veracity—during the first, second, or third answer of the sender.

In addition to examining the effect of the decision-making moment on credibility judgments, accuracy, and confidence, we sought to replicate several typical findings of the deception literature. First, as we were interested in examining how the moment influenced some phenomena such as the truth bias, the veracity effect, and so on (see a description of these phenomena below), we had to examine in the first place whether these effects were apparent in this study. Second, we also tried to replicate the typical findings because most deception research has been conducted in the US and the UK, and cultural differences may advise against generalizing these findings to other sociocultural settings (see Bond, Omar, Mahmoud, & Bonser, 1990; Cody, Lee, & Chao, 1989; Ruby & Brigham, 1997; Vrij & Winkel, 1991).

The specific hypotheses tested are explained in the following sections.

ACCURACY, TRUTH BIAS, AND VERACITY EFFECT

As stated above, extant research has found that overall detection accuracy (i.e., accuracy in detecting both truthful and deceptive communications; see Miller & Stiff, 1993) falls within the 45–60% correct judgments range, either not significantly different from chance, or only slightly greater than chance. However, before asserting that, generally, accuracy in deception-detection experiments is poor, a distinction must be made between accuracy in detecting truthful statements and accuracy in detecting deceptive statements (see Masip, Garrido, & Herrero, 2004). Levine, Park, and McCornack (1999) observed that research demonstrates that people generally tend to lend credibility to the verbal and nonverbal messages conveyed by others; that is to say, they show a *truth bias* (see, e.g., the reviews by Bond & DePaulo, in press; DePaulo, Stone, & Lassiter, 1985; Kalbfleisch, 1985; Vrij, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). Levine et al. also argued that this truth bias may give rise to a *veracity effect*, that is, accuracy in judging truthful statements will be significantly greater than accuracy in judging deceptive statements. In a series of three experiments, they repeatedly found that truth-detection accuracy was significantly greater than lie-detection accuracy (veracity effect). Also, accuracy in detecting the truthful statements was significantly greater than chance, while accuracy in detecting the deceptive statements was significantly lower than chance. This veracity effect was caused by a truth bias among observers. Indeed, the proportion of judgments of truthfulness was significantly greater than the actual distribution of truths and lies (truth bias), and a significant positive correlation was found between the proportion of judgments of truthfulness (hereafter PJT) and accuracy at detecting truthful statements, while a significant negative correlation was found between the PJT and accuracy at detecting deceptive statements. Levine et al. concluded that, since the overall accuracy rate blurs the differences between the detection of truthful and deceptive statements, deception researchers should report two separate accuracy measures, one for the truthful statements and one for the deceptive ones.

In view of the above arguments and findings of Levine et al. (1999), we posed the following hypotheses:

- *H1:* The observers will display a veracity effect:
  H1a: Accuracy in judging truthful statements will be significantly greater than accuracy in judging deceptive statements.
  H1b: Accuracy in judging truthful statements will be significantly greater than

chance, while accuracy in judging deceptive statements will be significantly lower than chance.

- *H2:* Observers will also show a truth bias: The frequency of judgments of truthfulness will be significantly greater than the frequency of judgments of deceptiveness.
- *H3:* This truth bias will account for the veracity effect. Thus, a significant positive correlation will be found between the PJT and accuracy in detecting truthful statements, and a significant negative correlation will be found between the PJT and accuracy in detecting deceptive statements.

The effects of the moment observers make their decision on the PJT and accuracy

Although these results should be expected generally, the moment observers make their decision might have an impact on the PJT and accuracy. There are at least two alternative ways this might happen, thus, two competing sets of predictions were made. First, from a systematic vs heuristic information processing framework, the PJT could be expected to decrease over time, particularly for the deceptive statements, while accuracy could be expected to increase over time.

These predictions are based on certain attribution and person perception theories, such as Fiske and Neuberg's (1990), Gilbert's (1989; Gilbert & Malone, 1995) or Trope's (1986). A common characteristic of these theoretical models is that they comprise several stages, the first of which is automatic, producing an immediate first impression based on a heuristic mode of processing. Only in a subsequent stage can the perceiver adjust his or her initial impressions. This requires a systematic way of processing information, which is effortful and requires cognitive resources (Chaiken, Liberman, & Eagly, 1989; Chen & Chaiken, 1999).

In the deception-detection domain, if credibility judgments are made very early (or if the behavioural segments to be judged are too brief), then observers are in the first, automatic step of these attribution and person-perception models. Also, observers may lack information to be processed in a systematic way in order to make a reasoned judgment, because the behavioural sample is so brief that it contains very little information. Therefore, the observers, requested by the experimenter to make a prompt credibility judgment, will make a heuristic judgment. And, normally, heuristic credibility judgments are judgments of truthfulness. For example, Stiff, Kim, and Ramesh (1992) conceptualized the truth bias as a cognitive heuristic. Also, Millar and Millar (1997) found a more pronounced truth bias under conditions of low cognitive capacity than under conditions of high cognitive capacity. Finally, Gilbert, Krull, and Malone (1990) found that interrupting the participants' processing of information increased the likelihood of them considering false linguistic or iconic propositions to be true, but not the likelihood of them considering true propositions to be false. The authors concluded that merely comprehending a proposition increases the likelihood of initially representing it as true, even if the proposition is actually false. Only in a further stage are respondents able to critically assess its veracity. All these findings suggest that initial credibility judgments should be judgments of truthfulness.

However, as time goes by, the PJT should decrease. First, because observers will reach a later stage in their perception process, so that the automatic initial impression can be changed. Second, because the growing amount of information will allow observers to make a reasoned judgment. If, in addition, it is assumed that, when processing the senders' information in a systematic way, the observers are somewhat capable of discriminating between truthful and deceptive testimonies,[2] then the decrease in the PJT will be much more pronounced in judging deceptive statements than in judging truthful statements, since the correct identification of the latter would lead observers to make judgments of truthfulness.

These processes should have an impact on accuracy. If the PJT decreases over time, especially in judging the deceptive statements, then accuracy in judging the deceptive statements will increase. Also, if, when judging truthful accounts, the PJT decreases slightly, does not decrease, or even increases, then truth-detection accuracy will also

[2]This assumption might seem questionable, since extant research indicates that human observers' accuracy in judging the senders' credibility falls within the 45–60% range. However, as stated in the text, these conclusions are largely based on research using very brief behavioural samples, forcing observers to make their judgments heuristically and preventing them from using enough information to make a more accurate judgment. Even in the studies in which longer behavioural samples have been used, it is not clear whether the observers waited until the end of the senders' performance to make their judgment. The common assumption that observers' judgmental accuracy is poor may be an artifact based on how deception research has been conducted so far.

decrease slightly, will remain the same, or will increase.

Thus, we posed the following hypotheses:

- *H4a:* The PJT will decrease over time, particularly for the deceptive statements.
- *H5a:* Accuracy in judging deceptive statements will increase over time.

The alternative predictions were based on Masip, Garrido, and Herrero's (2003b) report. Masip et al. explicitly examined the effect of the moment at which observers made their decision about the sender's credibility on observers' accuracy in detecting truthful and deceptive accounts. They argued that, since the information conveyed by the sender when telling the truth is accurate, the more information given, the more correct the judgments. This would yield an increase in the PJT and, hence, an increase in accuracy in judging truthful statements. Conversely, the growing information conveyed by the sender while lying is misleading, since the sender manipulates this information in order to create an apparently truthful statement (information management), and tries to convey the impression of him- or herself being honest (image and behaviour management) (Buller & Burgoon, 1996, 1998; Greene, O'Hair, Cody, & Yen, 1985; Vrij, 2000; Zuckerman et al., 1981). This would also result in an increase in the PJT and a decrease in accuracy in judging deceptive statements.

In a series of two experiments, Masip et al. (2003b) found partial support for their predictions. However, in their studies, Masip et al. used only one sender. This seriously questions the generalizability of their findings. Another limitation of Masip et al.'s experiments was the subjective nature of how the observers distinguished between the diverse moments. Masip et al. asked their observers to indicate when they had come to the conclusion that the sender was lying or telling the truth, at the beginning of the statement (Moment 1), at the middle part of it (Moment 2), or at the final moment (Moment 3). However, there were no clear boundaries between the critical time periods. Thus, what for some observers was Moment 1 may have been Moment 2 for others, and what for some observers was Moment 2 may have been Moment 3 for others.

In any case, based on Masip et al.'s considerations and findings we posed the following alternative hypotheses:

- *H4b:* The PJT will increase over time, for both the truthful and the deceptive statements.
- *H5b:* Accuracy will increase over time for the truthful statements, but will decrease for the deceptive ones.

CONFIDENCE

Research into respondents' accuracy in detecting truthful and deceptive messages has repeatedly examined observers' confidence in their credibility judgments. A meta-analysis of this literature concluded that, typically, respondents: (a) are unaware of how well they are doing in judging the credibility of truthful and deceptive messages; (b) display significantly more confidence in judging truthful statements than deceptive statements; and (c) are more confident in their judgments of truthfulness than in their judgments of deceptiveness (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997).

In line with these findings we predicted that:

- *H6:* Confidence in judging truthful accounts will be significantly higher than confidence in judging deceptive accounts.
- *H7:* Observers' confidence in their truthfulness judgments will be significantly greater than their confidence in their deceptiveness judgments.

The effects of the moment at which observers make their decision on observers' confidence

Few researchers have examined whether observers' confidence in their credibility judgments changes over time. Granhag and Strömwall (2001) conducted a study in which senders were interviewed three times about the same facts. One sample of observers had to judge the senders' credibility twice: first after watching the first interview, and then after watching the remaining two interviews. Granhag and Strömwall found that observers placed more confidence in their second judgment than in their first judgment. This might be due to the greater amount of information available to observers when they made the second judgment, or to them having a chance to test their initial hypotheses about the truthfulness of the first statement. However, unlike the observers in our study, Granhag and Strömwall's observers made two separate judgments about the same statements.

The prediction for the present study, in which observers have to make only one judgment per sender, is unclear. On the one hand, those who "are certain of it" from the very beginning might make their decision about the sender's credibility

earlier, since they feel there is no reason to wait. Thus, high confidence scores may be expected at Moment 1. However, those who have doubts about the senders' veracity will wait until they are sure enough to make a judgment. They can feel confident at Moment 2 or at Moment 3, in which case they will make their judgments at these moments. In this case, confidence would not change over time. Also, it is possible that Moment 3 arrives, observers must make a judgment, and they are still uncertain about the senders' veracity. Therefore, they will score low in confidence. In this case, confidence would decrease over time. In view of these conflicting alternatives no specific hypothesis was posed about the influence of the moment of decision on observers' confidence, but the issue was examined.

In order to test the above hypotheses, a study was conducted. The study was similar to that of Masip et al. (2003b), but the weaknesses of their design were addressed. Thus, in the present experiment: (a) 24 senders (not just 1) lied and told the truth about the facts they had witnessed on a videotape; and (b) these senders were asked three questions so that each observer's answer to the first question could be considered as Moment 1, the answer to the second question as Moment 2, and the answer to the third question as Moment 3. Thus, clear and unambiguous "markers" were used to differentiate between the three time periods of interest.

METHOD

Stimuli collection

Participants

Twenty-four undergraduate students of criminology at a Spanish University (13 females and 11 males; mean age 20.89 years, age range 18–29 years) volunteered to participate as senders in the study.

Procedure

Creating the videotapes to be shown to the senders. In the present experiment, each sender watched a videotape in which three characters appeared. One of these characters committed an offence. After watching the videotape, the senders were asked three questions: One about each of the three characters in the tape. They had to lie and tell the truth in response to these questions while being videorecorded. The tapes containing the senders' statements were later edited and shown to observers who had to judge whether each sender was lying or telling the truth. Each observer watched 12 three-answer statements. Six of such statements were truthful, while the remaining six were deceptive.

If all six truthful statements to be watched by any particular observer were based on the same facts, then the same statement would be repeated six times (i.e., six different senders would tell exactly the same story). As a result, towards the end of the 12-statement series, observers would know that those statements that were being repeated once and again (up to six times) were probably truthful. Therefore, instead of showing only one videotape to senders, we decided to show them two. This would decrease the number of repeated truthful statements to be watched by any particular observer from six to three. However, at the same time, certain characteristics of these tapes had to be kept constant so as not to influence the dependent measures. More specifically: (a) the characters had to be the same; (b) the offence and the general facts, as well as the physical setting, also had to be the same; and (c) each character had to perform different actions in each videotape.

Two scripts that met these conditions were written by the first author. The characters were a woman, a man with a moustache, and a man wearing a suit. The woman and the man with the moustache came into a waiting room. The second to come, who was carrying a briefcase and looked nervous, asked the other visitor the time. After a short while, the man wearing a suit came in and invited the first visitor to enter his office. The second one asked whether s/he could go in first, since s/he was in a hurry because s/he had to make a payment and the shops were about to close. S/he added, pointing to the briefcase, that s/he had just got the money from the bank. The man in the suit told them that they could go in together. Once inside his office, the man in the suit and the second visitor arranged something to do with a document. After that, the second visitor left, leaving his/her briefcase full of money on the office desk. Then, the first visitor, when the man in the suit was not looking, seized the briefcase and asked whether s/he could leave for a minute to check whether s/he had shut the door of his/her car.

The roles of the woman and the man with the moustache were exchanged in the two versions of the tape. Also, the actions performed by the man in a suit and the second visitor were different in each videotape. In addition, in one tape the man in a suit accompanied the second visitor to the door of his office to say goodbye, while in the other tape

he said goodbye from behind his desk. Finally, in one tape the man wearing a suit gave permission to the first visitor to leave, while in the other tape he realized that the first visitor had taken the second visitor's briefcase and adopted a harsh attitude towards him.

Three semiprofessional actors played the role of the three characters. Their performances were videorecorded. The setting was the waiting room and the office of the Dean of the Faculty of Psychology of the authors' university. The videos were recorded on a Saturday evening, when the building was closed and no one would interrupt the recording session. Later on, a technician edited the videotapes. He was asked to use the same shots in both videotapes, since they had to be as similar as possible except for the critical differences. The final tapes lasted over 3 minutes.

Interviewing senders. Criminology students at the Faculty of Law were recruited to participate as senders in the experiment. In order to motivate them, the senders were challenged to lie convincingly. They were told that after observers had made their judgments, a list with their names ranked according to their lying ability would be posted on a board. They were encouraged to lie convincingly in order to appear at the top of the list.

Twenty-four students volunteered and signed informed consent forms. They were cited to go to the Faculty of Psychology one at a time. Each sender was received by a research assistant in the faculty hall and was escorted to a room where a second research assistant told him or her that s/he had to watch a videotape and that, after 10 minutes during which s/he could prepare his or her answers, s/he would be asked some questions about the facts depicted in the tape.

Then the videotape was shown to the sender. Afterwards, the second research assistant told him or her that she was going to bring him or her to another room where an interviewer would ask some questions about what each character in the videotape did. She told the sender that the interviewer would ask those questions twice. In one case s/he had to tell the truth, while in the other case s/he had to lie. The research assistant announced that she was leaving for 10 minutes, so that s/he (the sender) could prepare his or her statements. This was done because in actual situations, such as police interviews of witnesses, the interviewee is not questioned immediately, but after some time, and can anticipate the kind of questions that will be made and plan in advance the answers that s/he will provide.

She left the room, and after 10 minutes the first research assistant came in and accompanied the sender to another room. The first author, who acted as the interviewer, received the sender and asked him or her to sit down on a chair located in front of two videocameras, one of which recorded the sender's face while the other recorded his or her full body. Only the full-body recordings were used in the experiment described in the present report. The cameras were placed behind the interviewer's chair in such a way that the interviewer did not appear on the videotapes.

In order to create a rapport with the sender, as well as to allow him or her to relax—since the presence of the two videocameras was somewhat intimidating—the interviewer and the interviewee spoke for about 10 minutes about informal topics unrelated to the research while the cameras were already functioning. Then, the interviewer told the interviewee that he was going to ask three questions and that each question would deal with the actions of one of the three characters in the videotape. He also announced that he was going to ask the questions twice, and that the sender was expected to tell the truth (to lie in the alternative condition) the first time (first interview) and to lie (to tell the truth in the alternative condition) the second time the three questions were asked (second interview). At this point, in order to increase his or her motivation, the sender was reminded of the contest and the liars' ranking, and the need to make an effort to lie convincingly was stressed.

All the questions in the interview had the same structure: ''Describe in detail what the man with a moustache (the man in a suit / the woman) did; I remind you that you have to tell the truth (lie).'' The interviewer was blind with regard to the tape that that individual sender had watched. After both interviews were completed, the interviewer invited the sender to ask any questions about the procedure, answered those questions, debriefed the sender, and thanked him or her for participating.

A series of variables were counterbalanced. Thus, while one half of the senders watched the first version of the videotape, the other half watched the second version. In each case, the order of the three questions was also counterbalanced, so that each question was in the same location (1st, 2nd or 3rd place) within the question sequence the same number of times. Whether senders were first requested to lie and then to tell the truth or vice-versa was also counterbalanced. Finally, an effort was made to have a male and a female sender within each cell of the video version × question sequence × lying-truthtelling order matrix. The study design is summarized in Table 1.

TABLE 1
Study design and video composition (statements in each video and their order)

Version of senders' videotape	*Sequence of questions*[a]	*Interview order*[b]	*Observers' tapes*[c]				*Statement order in observers' tapes*
			A1	*A2*	*B1*	*B2*	
Version 1	1-2-3	T-D	1D	2T	1T	2D	2nd
		D-T	3T	4D	3D	4T	12th
	2-3-1	T-D	5D	6T	5T	6D	9th
		D-T	7T	8D	7D	8T	10th
	3-1-2	T-D	9D	10T	9T	10D	6th
		D-T	11T	12D	11D	12T	4th
Version 2	1-2-3	T-D	13D	14T	13T	14D	7th
		D-T	15T	16D	15D	16T	3rd
	2-3-1	T-D	17D	18T	17T	18D	8th
		D-T	19T	20D	19D	20T	5th
	3-1-2	T-D	21D	22T	21T	22D	1st
		D-T	23T	24D	23D	24T	11th

[a]1: "Describe in detail what the man with a moustache did"; 2: "Describe in detail what the woman did"; 3: "Describe in detail what the man in a suit did."
[b]D-T: deceptive-truthful, that is, the sender lied in the first interview and told the truth in the second interview; T-D: truthful-deceptive, that is, the sender told the truth in the first interview and lied in the second interview.
[c]A1, A2, B1, B2: see text; 1D: deceptive statement of Sender 1; 2T: truthful statement of Sender 2; etc.

For example, in the first row we can see that Senders 1 and 2 (statements 1D [sender 1's deceptive statement], 1T, 2D, 2T) had watched the first version of the videotape (Version 1), were asked the questions in the 1-2-3 sequence, and first told the truth and then lied (T-D).[3]

Video editing. The 48 statements that had been collected (a three-answer truthful and a three-answer deceptive statement from each sender; the interviewer's questions were removed) were edited to build four different 12-statement videotapes. The truth value of the statements, the version of the original videotapes, the order of the questions asked by the interviewer, and whether the senders had first lied and then told the truth or vice-versa, were counterbalanced in the tapes. In addition, the same sender never appeared in the same tape lying and telling the truth.

The resulting videotapes were labelled A1, A2, B1, and B2 (see Table 1). A random series of numbers ranging from 1 to 12 determined the order of the statements within each videotape. Thus, as shown in Table 1 (column on the right), the first statement contained in Videotape A1 was the deceptive account of sender 21 (21D), the second statement was the deceptive account of sender 1 (1D), and so on.

As stated above, each statement contained the truthful or deceptive answers to three questions. The questions had been removed from the videotapes that had to be shown to observers. On average, the senders' answer to the first question had a duration of 51 s and their answers to the second and third questions both had a duration of 52 s. In the stimulus videotapes, each series of three answers given by any particular sender was preceded by a 2-s dark screen with the heading "Subject X," where "X" was a number between 1 (the first sender in the tape) and 12 (the last sender in the tape). Within each statement, another 2-s dark screen separated the sender's different answers.

[3]It is apparent that the interviewer was not blind to the veracity of the senders' statements. This could potentially have influenced the way he interacted with each sender. While this is a frequent risk in much deception research, a number of safeguards were taken in this experiment to prevent this from happen. The interviewer was more an experimenter giving instructions to the sender than a real interviewer. In other words, the substantial part of the interview was not a free interaction, but was tightly scripted, and the wording of the questions was always the same. In no case did the interviewer interrupt the interviewee to make questions or comments. In addition, before the actual data collection began, the interviewer rehearsed his performance while interviewing mock participants in the same room where the actual interviews were to be conducted. Several research assistants were present in these rehearsals and provided feedback as to the verbal and nonverbal behaviour of the interviewer, which had to be neutral.

Data collection

Participants

The observers were 54 undergraduate students of psychology at a Spanish University (46 females and 8 males; mean age 22.41 years, age range 21–29 years). They received class credit for their participation.

Procedure

Observers were allocated to one of four separate groups. Each group had to judge a different videotape. Allocation of observers to the different groups was made randomly, except that the few male respondents were evenly distributed across groups. All respondents within a group were asked to attend the experimental session together. After they had entered a room equipped with a video and a TV monitor, an answer booklet was given to them. The instructions were printed on the front page of the booklet and the experimenter gave them orally as well. The observers were told that they were about to watch a series of 12 people making statements. Each person had watched a videotape and had been asked three questions about the events depicted in the videotape. They were informed that the questions had been suppressed from the tape. Also, they were told that the statement of each particular sender could be truthful or deceptive. It was stressed that it was not possible for a sender to be either telling the truth or lying in only one or two of his or her answers. Rather, either all three answers of any particular sender were truthful, or all three answers were deceptive. The observers were asked to determine whether each sender was lying or telling the truth. After they had watched the three answers of each sender, the experimenter would pause the tape in order for them to have time to express their judgments by ticking the "Lying" or "Telling the truth" box of this particular sender in the response booklet. Observers were also asked to indicate their confidence in each judgment on a scale ranging from 1 (*not at all confident*) to 7 (*absolutely confident*). Also, they had to indicate, for each judgment, whether they had made their decision concerning that sender's veracity during his or her first, second, or third answer.

Also, since there were still three truthful statements on each tape that were based on the same version of the videotape shown to the senders, it was important to prevent the observers from making cross-statement comparisons, because this could artificially increase their later judgments. Therefore, they were told that: (a) not all senders had watched the same tapes; (b) on most occasions, but not always, the videotapes that the senders had watched were very similar; and (c) on most occasions, but not always, deceivers did not drastically alter the facts depicted in the videotape they had watched, but introduced minor changes. Thus, as a result, truthful and deceptive statements were generally very similar. The observers were then advised to judge each statement independently from the others.[4] Also, the observers were told that there were three characters in all the videotapes watched by the senders: a woman and two men. They were also informed about the questions that the senders had been asked.[5]

After the observers had been given and had understood all these instructions, the experimental session began. The experimenter, placed behind the observers, used the remote control to pause the videotape after each statement, and did not press the "play" button again until all the participants had answered all the questions in the response booklet about this particular statement. After the observers had completed the task for all 12 statements, the experimenter thanked them for their participation, answered their questions, asked them not to talk about the study to those participants who had to act as observers later on,

[4]In order to examine whether accuracy for later judgments increased, manipulation check analyses were conducted. There were two sets of senders: odd senders and even senders. Odd senders were in videos A1 and B1, whereas even senders were in videos A2 and B2. The order of senders in A1 was the same as in B1, but those senders who lied in A1 told the truth in B1, and vice-versa (Table 1). The same can be said of A2 and B2. All videotapes contained 12 statements. Accuracy in judging the last 6 statements (which contained 3 truths and 3 lies) was compared with accuracy in judging the first 6 statements (also 3 truths and 3 lies) for both odd and even senders. These comparisons were not significant, $t(26) = 1.40$, $p = .173$, for odd senders, and $t(26) = 1.68$, $p = .104$ for even senders.

[5]The experimental instructions were tested on three samples. The convenience of including all this additional information was assessed. It became apparent from the comments of the participants in the pilot tests (undergraduate students who received the instructions and performed part of the task) that, unless they received this information, they assumed that all senders had watched the same videotape and then tended to compare different statements in order to assess their veracity. Furthermore, it was certain that, except for a few cases, the observers did not make major changes when lying. Also, since the questions had been removed from the tapes, we thought that telling observers what the questions were about would help them to focus upon the experimental task—instead of, for instance, trying to guess the interviewer's question. The pilot test supported our views. Observers in the actual experimental task were also informed that the three questions had been counterbalanced.

and informed them that they would be invited to attend a lecture where they would be debriefed.

RESULTS

Accuracy and proportion of judgments of truthfulness (PJT)

Accuracy

Overall accuracy (proportion of correct judgments) was very moderate (.54) but significantly greater than chance probability (.50), $t(51) = 2.50$, $p = .016$. As predicted in Hypothesis 1, accuracy in judging truthful accounts (.59) was significantly greater than chance, $t(51) = 3.04$, $p = .004$. However, contrary to our prediction, observers' accuracy in judging deceptive statements (.50) did not differ from chance, $t(51) = -0.12$, $p = .907$.

These results suggest that accuracy was greater when judging truthful statements than when judging deceptive statements. To examine whether the difference was significant, an analysis of variance (ANOVA) was conducted on the proportion of correct judgments. The independent variables were the value of truth of the statements (truthful vs. deceptive statement) and the stimulus videotape (A1/A2/B1/B2), with repeated measures in the first variable. A significant main effect for value of truth revealed that accuracy in judging truthful accounts was significantly greater than accuracy in judging deceptive accounts, $F(1, 50) = 4.80$, $p = .033$. However, as shown in Figure 1, this main effect was qualified by a significant Value of Truth × Videotape interaction, $F(3, 50) = 4.27$, $p = .009$. This interaction indicated that the strength of the veracity effect was not the same across all four videotapes—in fact, a nonsignificant trend, $t(12) = -0.92$, $p = .376$, in the opposite direction was found for B2. The main effect for the videotape was not significant in this ANOVA, $F(3, 50) < 1$.

The correlation between accuracy in judging truthful statements and accuracy in judging deceptive statements was negative and did not reach statistical significance, $r = -.23$, $p = .092$. As might be expected, both accuracy in judging truthful statements and accuracy in judging deceptive statements correlated significantly with overall accuracy, respectively, $r = .68$ and $r = .56$, both ps $< .001$.

Proportion of judgments of truthfulness

In support of Hypothesis 2, more judgments of truthfulness were made than judgments of deceptiveness: Across truthful and deceptive statements

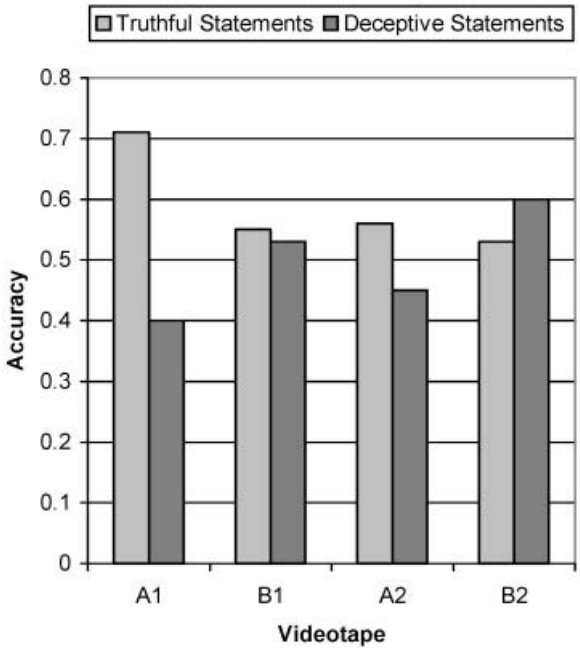

Figure 1. Accuracy rates.

the proportion of judgments of truthfulness (.55) was significantly greater than chance, $t(51) = 2.12$, $p = .039$. The PJT in judging truthful statements (.59) was also greater than chance, $t(51) = 3.04$, $p = .004$, but, in judging the deceptive statements, the PJT (.50) did not differ from chance probability, $t(51) = 0.12$, $p = .907$.

These results suggest that the PJT was greater when judging the truthful statements than when judging deceptive ones. An ANOVA was conducted on the PJT. The independent variables were the value of truth of the statements (truthful vs. deceptive statement) and the stimulus videotape (A1/A2/B1/B2), with repeated measures in the first variable. A significant main effect for the value of truth on the PJT indicated that significantly more judgments of truthfulness were made when the statements were actually truthful (.59) than when they were deceptive (.50), $F(1, 50) = 6.19$, $p = .016$ (Figure 2). This suggests that observers were somewhat capable of discerning between truthful and deceptive accounts, as already indicated by the accuracy results reported above (overall accuracy was greater than chance). The main effect for the videotape was also significant, $F(1, 50) = 4.27$, $p = .009$. In general, the statements contained in some videotapes were more often considered truthful than those contained in other videotapes (see Figure 2). Indeed, this effect is consistent with the above finding of a significant Truth Value × Videotape interaction on accuracy: The greater the PJT, the greater the proportion of accurate judgments in judging

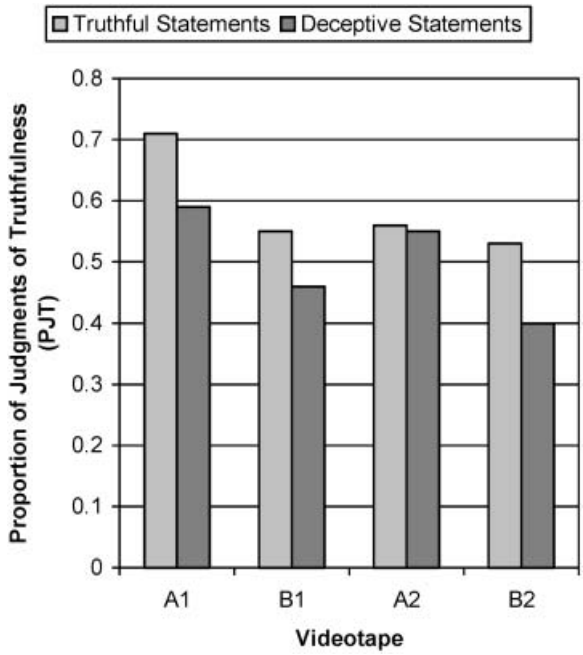

Figure 2. Proportion of judgments of truthfulness (PJT)

truthful statements and the smaller the proportion of accurate judgments in judging deceptive statements. Greater veracity effects were caused by greater PJTs.

Interestingly, the Value of Truth × Videotape interaction was far from significance, $F(3, 50) < 1$. This indicates that, despite the differences among the videotapes in terms of the veracity effect, the above main effect for the value of truth on the PJT was consistent across all the videotapes that were shown to observers.

As predicted in Hypothesis 3, the PJT across truthful and deceptive statements bore a significant positive correlation with accuracy in judging truthful statements, $r = .81$, $p < .001$, and a significant negative correlation with accuracy in judging deceptive statements, $r = -.76$, $p < .001$. As might be expected, the correlation between the PJT and overall accuracy was not significant, $r = .12$, $p = .395$. The correlation between the PJT in judging truthful statements and the PJT in judging deceptive statements did not reach statistical significance, $r = .23$, $p = .092$. Both the PJT in judging truthful statements and the PJT in judging deceptive statements correlated significantly with the overall PJT, respectively, $r = .81$ and $r = .76$, both ps $< .001$.

The influence of the moment at which observers made their decision

Since many participants, either in judging the truthful statements or in judging the deceptive ones, did not choose to make any decision at Moment 1, at Moment 2, or at Moment 3, the moment variable could not be entered as an independent variable in the above ANOVAs. Instead, its effect on accuracy, the PJT, and confidence, was examined using correlations. Moment-1 judgments were coded as 1, Moment-2 judgments as 2, and Moment-3 judgments as 3. Then, these values were summed across truthful statements, across deceptive statements, and across both truthful and deceptive statements. Positive correlations between these variables and accuracy, the PJT, and confidence, would reflect increases in accuracy, PJT, and confidence as observers made their decisions later. Negative correlations would reflect decreases in accuracy, PJT and confidence.[6]

Accuracy. Overall accuracy did not change significantly over time, $r = -.01$, $p = .933$. For truthful statements, the moment–accuracy correlation was negative but nonsignificant, $r = -.14$, $p = .309$. For deceptive statements, this correlation was positive and marginally significant, $r = .24$, $p = .077$, revealing a trend for observers to make more accurate judgments of the deceptive statements the later they made their decisions.

Proportion of judgments of truthfulness (PJT). A statistically significant moment–PJT correlation indicated that the PJT decreased when observers made their decisions later, $r = -.29$, $p = .037$. While this reduction was very meagre for the truthful statements, $r = -.14$, $p = .309$, it reached a marginal significance for the deceptive ones, $r = -.24$, $p = .077$.

Moment in judging truthful and deceptive statements. The correlation of the moment variable for truthful statements with the moment variable for deceptive statements was positive and significant, $r = .34$, $p = .013$. This indicated that, regardless of the truth value of the statements, some observers consistently made their decisions earlier than others. As might be expected, the moment variable, when collapsed across both truthful and deceptive statements, was significantly correlated

[6]In order to conduct Pearson correlations the distributions must be normal. The skewness and kurtosis of accuracy, the PJT, confidence, and the moment variable (for truthful statements, deceptive statements, and across all statements) were calculated using the SPSS "frequencies" command and were then divided by their standard error. Most of the resulting values were lower than 1, and all of them were lower than 2. Therefore, the distributions did not depart significantly from normality.

with the moment variable for truthful statements, $r = .79$, $p < .001$, and deceptive statements, $r = .84$, $p < .001$.

Confidence

Descriptive results. Confidence across both truthful and deceptive statements was 4.18 on a 1–7 scale. Rates were 4.22 for the truthful statements and 4.15 for the deceptive statements. An ANOVA similar to those conducted on accuracy and the PJT was run taking the observers' judgmental confidence as the dependent measure. Contrary to Hypothesis 6, the truth value of the statements had no significant effect on confidence, $F(1, 50) < 1$. Similarly, neither the stimulus videotape, $F(3, 50) < 1$, nor its interaction with the value of truth of the statements, $F(3, 50) = 2.04$, $p = .121$, had any significant effect on confidence.

Confidence–accuracy correlations. In line with the results of DePaulo et al. (1997), the relationship between confidence and accuracy was not significant, $r = .13$, $p = .347$. When examining this relationship separately for truthful and deceptive statements, it was apparent that confidence and accuracy in judging deceptive statements were unrelated, $r = -.20$, $p = .156$. However, a significant correlation emerged between confidence and accuracy in judging truthful statements, $r = .32$, $p = .019$.

The influence of the moment at which observers made their decision. The correlation between the moment at which observers made their decision and their confidence in that decision was negative and marginally significant, $r = -.25$, $p = .066$. This effect was negligible for the truthful statements, $r = -.15$, $p = .280$, but was significant for the deceptive ones, $r = -.33$, $p = .015$. The later observers made their decision, the less confidence they had in that decision.

PJT–confidence correlations. Both the above significant correlation between confidence and accuracy in judging truthful accounts and the decrease in confidence over time when judging deceptive accounts could be accounted for by the relation predicted in Hypothesis 7: Greater confidence in judgments of truthfulness than in judgments of deceptiveness. First, if confidence in making judgments of truthfulness is particularly strong, the more judgments of truthfulness observers make, the greater their mean confidence scores. If the statements are truthful, accuracy will rise as well. This would account for the above significant correlation between confidence and accuracy for the truthful statements. In support of this explanation, the PJT–confidence correlation was positive and significant for the truthful statements, $r = .32$, $p = .019$. Second, when judging the deceptive statements, when observers make their decision later not only their confidence decreases, but also the PJT. If greater confidence is placed in the judgments of truthfulness than in those of deceptiveness, then as the PJT decreases over time, so will the observers' confidence. However, for the deceptive statements, the PJT–confidence correlation was not significant, $r = .20$, $p = .156$. Finally, across both truthful and deceptive statements, the PJT–confidence correlation did not reach statistical significance, $r = .23$, $p = .092$.

Confidence in judging truthful and deceptive statements. Those observers who were more confident in judging truthful statements were also more confident in judging deceptive statements, $r = .74$, $p < .001$. This, coupled with the above findings that observers' confidence was unaffected by the statement truth-value, the stimulus videotape, or their interaction, suggests that some observers were more confident than others. As might be expected, overall confidence across truthful and deceptive statements was significantly related to both confidence in judging truthful statements, $r = .94$, $p < .001$, and confidence in judging deceptive statements, $r = .93$, $p < .001$.

DISCUSSION

Accuracy, truth bias, and veracity effect

Deception research has normally been conducted using very brief behavioural segments, and has led researchers to conclude that observers making credibility judgments show a strong truth bias. In this experiment, we examined how the length of the behavioural segments influenced the judgments, accuracy, and confidence of the observers. Also, we sought to replicate several typical findings of the US and North-European deception literature in a South-European country. The results indicate that, in line with the findings of extant research conducted in other countries (Bond & DePaulo, in press; DePaulo et al., 1985; Vrij, 2000), overall detection accuracy is poor but significantly greater than chance. However, when the truth value of statements was taken into

consideration, following Levine et al.'s (1999) recommendations, a more complex picture emerged. Thus, as predicted in Hypothesis 1a, accuracy in judging truthful statements was significantly greater than accuracy in judging deceptive statements. Therefore, a veracity effect was apparent. Furthermore, although accuracy in judging deceptive statements was not below chance, accuracy in judging truthful statements was significantly greater than chance (Hypothesis 1b). This veracity effect seemed to be caused by a truth bias. Indeed, supporting Hypothesis 2, the PJT was greater than chance—or, in other words, significantly more judgments of truthfulness were made than judgments of deceptiveness. In addition, the PJT bore a significant positive correlation with accuracy in judging truthful statements and a significant negative correlation with accuracy in judging deceptive statements (Hypothesis 3). Also, the correlation between the PJT in judging truthful and deceptive statements was not significant. These results are generally consistent with Levine et al.'s (1999) arguments, suggesting that overall accuracy rates are misleading, since large differences, caused by the truth bias and veracity effect phenomena, are apparent between accuracy in judging truthful and deceptive communications.

However, some further considerations are pertinent. First, as stated above, the PJT across both truthful and deceptive accounts was significantly greater than chance. However, consistent with an accuracy rate above chance, the PJT was significantly greater when judging the truthful statements than when judging the deceptive ones. In fact, while in the former case it was greater than chance, in the latter case it did not differ significantly from chance. These findings question the notion that the increased PJT was merely due to a truth "bias." Although when judging the deceptive statements the PJT was indeed excessive (it should have been significantly lower than chance), the overall PJT across truthful and deceptive statements was greater than chance because of the contribution of the PJT in judging truthful accounts. However, if observers are assumed to possess some detection skills, their PJT in judging truthful statements *must* be greater than chance, as well as greater than their PJT in judging deceptive statements. But this is not indicative of a truth bias, but rather of observers' accuracy in detecting truthful statements.[7]

Second, the moment when observers made their decisions correlated significantly with the PJT. As expected if initial credibility judgments are automatic while later judgments are made on the basis of a systematic processing of information, the PJT decreased over time. Furthermore, it decreased more for the deceptive statements than for the truthful ones—for which the reduction did not even approach significance. This reflects an increased ability among observers to progressively discriminate between truthful and deceptive accounts, as would be expected if they progressively engaged in systematic information processing. The nonreduction in the PJT in judging truthful statements does not reflect the existence of a truth "bias," since the accurate judgments of the truthful statements are indeed judgments of truthfulness. These results support Hypothesis 4a, and are contrary to Hypothesis 4b.

The changes in the PJT had an impact on accuracy. Since the PJT in judging truthful statements did not change over time, accuracy in these statements was similar at all three moments. Also, since the PJT in judging deceptive accounts decreased over time, accuracy in judging deceptive statements increased. However, the progressive differentiation between truthful and deceptive statements over time was not sufficient to increase overall detection accuracy (i.e., accuracy in detecting both truthful and deceptive accounts).

These results do not support Masip et al.'s (2003b) hypotheses predicting an increase in accuracy in judging the truthful accounts and a decrease in judging the deceptive ones. Instead, they point to the heuristic vs systematic information processing explanation. Previous research conducted by Granhag and Strömwall (2000, 2001) points in the same direction. They found a significant increase in accuracy when the same observers made two credibility judgments of the same senders. The increase was marginal for the deceptive statements and significant for the truthful ones. Although it was not a central point of their study and therefore they overlooked it, it is interesting that Granhag and Strömwall's results further support our heuristic vs systematic processing prediction: in their study, not only did the PJT

[7]An exaggerated number of judgments of truthfulness were only apparent when deceptive accounts were judged. Although the PJT when judging the truthful accounts (.59) was greater than chance, it was far from 1. As for the deceptive statements, the .50 PJT might reflect, as an anonymous reviewer suggested, that participants relied on mere guesswork. Alternatively, since the moment–PJT correlation was significant, it can be assumed that at Moment 1 the PJT in judging deceptive statements was greater than the average .50 rate, whereas at Moment 3 it was lower. Therefore, a truth bias was presumably present when judgments of deceptive statements were made early. Unfortunately, the design of this study made it difficult to calculate the PJT separately at Moments 1, 2, and 3. Ongoing research is addressing this issue.

decrease for the deceptive statements (although only marginally), but it also increased significantly for the truthful ones, thus reflecting an increased ability to distinguish between both kinds of statements. However, Granhag and Strömwall's results are not fully comparable with the present findings. For example, their temporal variable was a within-subject one, while our participants made just one judgment per sender. This might have influenced the results. Furthermore, Granhag and Strömwall explored the effects on accuracy of the *number of interviews* watched by observers, whereas we explored, *within the same interview*, the impact that the moment the decision about the senders' veracity was made had on the PJT and accuracy. Indeed, Granhag and Strömwall's second and third interviews did not consist of participants' answers to just one question, while our Moments 2 and 3 consisted of answers to single questions. Thus, presumably, Granhag and Strömwall's second and third interviews were longer and more informative than our participants' answers to the second and third questions. In addition, while Granhag and Strömwall's first interview was made 3 hours after having watched a staged event, the second interview 4 days later, and the last interview 1 week later, our first question was made a few minutes after the sender had watched the videotape, and the second and third questions were made immediately after that. Changes in memory over time, rehearsal time, and similar, are variables that might influence the senders' statement and the observers' judgments and accuracy. Despite these and other differences, the results of both studies are consistent with the heuristic vs systematic processing hypotheses. This strongly supports the validity of these hypotheses.

Confidence

In line with the results of DePaulo et al.'s (1997) meta-analysis, the correlation between confidence and accuracy was not significant when judgments made in assessing truthful and deceptive statements were taken together. Also, this correlation was not significant for the deceptive accounts. However, confidence and accuracy in judging truthful statements were significantly correlated. This was due to the positive relationship between confidence and the PJT predicted in Hypothesis 7: More confidence would be placed in judgments of truthfulness than in judgments of deceptiveness. This hypothesis was supported for the truthful statements. Then, since confidence in making truthfulness judgments was greater than confidence in making deceptiveness judgments, an increased number of judgments of truthfulness did not only increase observers' accuracy, but also their mean confidence. That is why confidence and accuracy in judging truthful statements bore a significant correlation. Hypothesis 7 was not, however, supported for the deceptive statements. Across both kinds of statements the confidence–PJT correlation only reached a marginal significance ($p = .092$).

Contrary to our sixth prediction, the truth value of statements did not influence observers' confidence. Although DePaulo et al. (1997) reported a significant positive correlation, the individual studies included in their meta-analysis varied notoriously, ranging from .60 (Waxter, 1983) to −.55 (Allen & Atkinson, 1981), and including .00 (Köhnken, 1987). Also, recent research has failed to replicate this correlation (Masip, Garrido, & Herrero, 2003a). Future research should examine what variables account for these discrepancies.

Two possibilities were raised about confidence over time. The first one did not anticipate any change; the second one predicted a decrease in confidence as observers decided later. The results are ambiguous. Confidence across truthful and deceptive statements decreased over time, but this decrease reached only a marginal significance ($p = .066$). Probably, there were observers who waited to make their judgments until they were confident enough and reached this confidence at Moments 2 or 3, while others did not reach enough confidence and made low-confidence judgments at Moment 3. Although this explanation can account for the marginal trend towards a reduction over time of the mean confidence across truthful and deceptive statements, it cannot explain why, while the decrease was negligible when judging the truthful statements, it was significant ($p = .015$) when judging the deceptive ones.

Conclusions

In summary, most research findings reported in the US and North-European literature were replicated. However, most of the results were under the influence of the moment at which observers made their judgmental decision. Consistent with the idea that initial credibility judgments are made heuristically while later judgments are made in a systematic manner, the PJT decreased over time, particularly for the deceptive statements. This yielded an increase in accuracy in judging false accounts. The truth bias detected in deception research may be caused by

researchers using behavioural samples that are too brief and uninformative. Finally, using the present design, the moment at which observers made their decision had only a marginal negative influence on confidence.

A limitation of this study is that participants were not randomly allocated to Moment 1, 2, or 3. Rather, they decided themselves when to make their judgments. As a result, individual differences may have influenced the results, since different people might have decided at different moments. Ongoing research is explicitly addressing this question.

Manuscript received December 2004
Revised manuscript accepted August 2005

REFERENCES

Allen, V. L., & Atkinson, M. L. (1981). Identification of spontaneous and deliberate behaviour. *Journal of Nonverbal Behaviour*, *5*, 224–237.

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

Bond, C. F., Jr, Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behaviour*, *14*, 189–204.

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

Buller, D. B., & Burgoon, J. K. (1998). Emotional expression in the deception process. In P. A. Andersen & L. Guerrero (Eds.), *Handbook of communication and emotion. Research, theory, applications and contexts* (pp. 381–402). San Diego, CA: Academic Press.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Borgh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.

Cody, M. J., Lee, W.-S., & Chao, E. Y. (1989). Telling lies: Correlates of deception among Chinese. In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology. An international perspective* (pp. 359–368). Amsterdam: Elsevier.

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*, 346–357.

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.

DePaulo, B. M., Zuckerman, M., & Rosenthal, R. (1980). Humans as lie detectors. *Journal of Communication*, *30*, 129–139.

Feldman, R. S., Jenkins, L., & Popoola, O. (1979). Detection of deception in adults and children via facial expressions. *Child Development*, *50*, 350–355.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, *72*, 1429–1439.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

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*, 601–613.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Granhag, P. A., & Strömwall, L. A. (2000). "Let's go over this again…" Effects of repeated interrogations on deception detection performance. In A. Czerederecka, T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.), *Forensic psychology and law. Traditional questions and new ideas* (pp. 191–196). Kraków: Institute of Forensic Research Publishers.

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection based on repeated interrogations. *Legal and Criminological Psychology*, *6*, 85–101.

Greene, J. O., O'Hair, H. D., Cody, M. J., & Yen, C. (1985). Planning and control of behaviour during deception. *Human Communication Research*, *11*, 335–364.

Kalbfleisch, P. J. (1985). Accuracy in deception detection: A quantitative review (Doctoral dissertation, Michigan State University, 1986). *Dissertation Abstracts International*, *46*, 4453B.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, *2*, 1–17.

Kraut, R. (1980). Humans as lie detectors. Some second thoughts. *Journal of Communication*, *30*, 209–216.

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.

Masip, J., Garrido, E., & Herrero, C. (2003a). Facial appearance and judgments of credibility: The effects of facial babyishness and age on statement credibility. *Genetic, Social, and General Psychology Monographs*, *129*, 269–311.

Masip, J., Garrido, E., & Herrero, C. (2003b). When did you conclude she was lying? The impact of the moment the decision about the sender's veracity is made and the sender's facial appearance on police officers' credibility judgments. *Journal of Credibility Assessment and Witness Psychology*, *4*, 1–36.

Masip, J., Garrido, E., & Herrero, C. (2004). The nonverbal approach to the detection of deception: Judgmental accuracy. *Psychology in Spain*, *8*, 48–59.

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

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

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin*, *29*, 1316–1327.

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.

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

Trope, J. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–257.

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

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behaviour: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behaviour*, *15*, 169–184.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of human communication*. New York: Norton.

Waxter, P. H. (1983). Emotional deceit: False words versus false action. *Motivation and Emotion*, *7*, 365–376.

Zuckerman, M., DeFrank, R. S., Hall, J. A., Larrace, D. T., & Rosenthal, R. (1979). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology*, *15*, 378–396.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14*, 1–59.