anales de psicología
2001, vol . 17, nº 1 (junio), 101-120

Murcia (España). ISSN: 0212-9728

Is There A Kernel Of Truth In Judgements Of Deceptiveness?

Jaume Masip* and Eugenio Garrido

University of Salamanca

Título: ¿Existe un granito de verdad en juicios sobre el engaño?

Resumen: Este estudio consiste en una réplica ampliada del trabajo de Bond, Berry y Omar (1994). Ochenta y cinco estudiantes universitarios cumplimentaron un formulario para indicar en qué experimentos, de una lista de siete estudios en que tendrían que mentir y dos en que tendrían que decir la verdad, estarían dispuestos a participar, así como en qué medida lo estarían. Asimismo, indicaron en escalas de siete puntos el grado en que personas que los conocían bien los consideraban veraces o mentirosos, y cuanto lo eran en realidad. Además, proporcionaron una fotografía tipo carnet de sus rostros. Estas fotografías se mostraron a dos muestras de observadores. La primera juzgó su honestidad y veracidad. La segunda evaluó su atractivo y aniñamiento facial. Los resultados muestran que no hubo relación entre las valoraciones de la honestidad o veracidad efectuadas por los observadores y la disposición de los participantes para colaborar en experimentos que implicaran engaño. Aunque las autoevaluaciones de sinceridad de los participantes no correlacionaron con su sinceridad real, sí lo hicieron las de quienes los conocían bien –proporcionadas por los propios participantes–. Ni las autoevaluaciones de los participantes sobre su sinceridad ni las de personas próximas a ellos se basaron en la apariencia facial. El atractivo físico y el aniñamiento facial estaban relacionados marginalmente entre sí, y no guardaban ninguna relación con la veracidad real ni percibida. La mayoría de los estudiantes estuvieron de acuerdo en participar en la mayor parte de los experimentos que implicaban engaño, y no expresaron fuertes cuestionamientos éticos contra el acto de mentir.

Palabras clave: Sinceridad, veracidad, engaño, mentira, rostro, cara, atractivo, aniñamiento facial.

Abstract: This study is an extended replication of Bond, Berry and Omar´s (1994) work. Eighty-five undergraduate students completed a form to indicate in which experiments, from a list of seven deceptive and two truthful studies, they would be willing to participate, as well as the strength of their willingness to participate in a 10-point scale. Also, they reported in 7-point scales to what extent people who knew them well thought they were truthful or deceptive, and how truthful or deceptive they were in reality. In addition, participants provided us with a passport-type photograph of their faces. These photographs were shown to two samples of observers. The first judged their honesty and truthfulness. The second assessed their attractiveness and babyishness. Results show there was no relation between observers´ honesty or truthfulness ratings and participants´ willingness to collaborate in deceptive experimental procedures. Although participants´ self-reported honesty did not correlate with their actual honesty, close acquaintances´ impressions, as reported by the participants, did. Neither participants´ self-reported honesty nor close acquaintances´ views were based on targets´ facial appearance. Attractiveness and babyfacedness were unrelated to real and perceived honesty, and were marginally related to each other. Most of the students agreed to participate in most of the deceptive experiments, and they expressed no strong ethical concerns against lying.

Key words: Honesty, truthfulness, deception, face, attractiveness, babyfacedness, babyshness, kernel of truth.

Throghout history, there seems to have been a popular belief in the relationship between physical appearance and personality. This has been so among both lay people and scientists alike. Gall and Spurzheim´s phrenology (e.g., Fancher, 1988), and Kretschmer (1921) and Sheldon´s (Sheldon, Stevens, & Tucker, 1940; Sheldon & Stevens, 1942) somatic typologies are but two examples of scientific attempts to lay the foundations of these popular beliefs about the existence of a bond between the body and the soul. Another similar "pseudopsychology" (Yates, 1967) is *physiognomy*: "the practice of trying to judge character and other psychological qualities by observation of facial features." (Alley, 1988, p. 167). There is evidence of physiognomic practices dating from ancient Greece and ancient China (Caro-Baroja, 1987; Zebrowitz, 1997), and the physiognomic discipline evolved through the Middle Ages (e.g., Avicena, Fakhr; see Viguera, 1977),

*__Address for correspondence__: 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@gugu.usal.es

the Italian Reinassaince (e.g., della Rocca, 1536; della Porta, 1644), and, thanks to the contributions of Lavater (1793; Moreau, 1820), Galton (1883) and Lombroso (1895), also through the 18th and 19th centuries (for a detailed history of physiognomy see Caro-Baroja, 1987).

In general, 20th century psychological research does not lend support to the often naïve claims made by the early physiognomists, although the general public still believes structural features of the human face strongly reflect one´s character. In this regard, Alley (1988) concludes from his review on the topic that "physiognomy is, with few and nearly negligible exceptions, an invalid practice, yet consistent facial stereotypes exist such that certain faces or facial characteristics produce remarkably uniform impressions in perceivers." (p. 185). Similar conclusions are reached by Bull and his colleagues in their reviews of research on the relationship between facial appearance and criminality (Bull, 1982; Bull & Green, 1982; Bull & McAlpine, 1998). This does not necessarily imply that personal information other than one´s psychological traits –information such as targets´ age, sex, race, identity, fitness, etc.– cannot be ascertained from facial static characteristics such as facial structure, skin colour or smoothness, etc. (e.g., Bruce & Young, 1998; Zebrowitz, 1997).

There are, however, a number of recent studies which show personality judgments based on targets´ facial appearance are accurate. This has been found for traits such as intelligence, extraversion, conscientiousness, agreeableness, emotional stability, dominance, etc. (e.g., Ambady & Rosenthal, 1992; Berry, 1990, 1991a; Borkenau & Liebler, 1993; Kenny, Albright, Malloy, & Kashy, 1994; Kenny, Horner, Kashy, & Chu, 1992; Levesque & Kenny, 1993; Moskowitz, 1990; Zebrowitz, 1997), giving rise to what has come to be known as *the kernel of truth hypothesis*. Yet, explanations provided by current researchers for these relationships between appearance and personality are more plausible and scientifically-sound than the weird or mystical ones posed by the ancient physiognomists. An outstanding example is Zebrowitz´s (1997) integrative model to account for appearance-trait relations. In this model, it is acknowledged that biological factors can influence both one´s physical appearance and one´s personality, as held by some of the historical authors (e.g., Lombroso, 1895; Sheldon *et al.*, 1940). However, the congruence or incongruence between facial appearance and personality may be due to other reasons as well, since the social and physical environment, psychological characteristics, and facial appearance influence each other through several basic psychological and psychosocial processes such as self-fulfilling and self-defeating prophecy effects, artifice effects, Dorian Gray effects, etc. (see Zebrowitz, 1997).

One personal characteristic that has been targeted by face researchers is honesty. The foci of interest have been two. First, researchers have been trying to ascertain whether some faces are more often than others judged to be honest, and the latter are more often than the former judged to be dishonest. In other words: are there socially shared stereotypes as to what honest or dishonest faces look like? Second, when observers agree as to which faces look honest and which look deceptive (*consensus at zero acquaintance*), are their perceptions accurate, that is, are stimulus individuals whose faces look dishonest really less honest than those whose faces look honest?

Experimental evidence clearly supports the first question –that some faces are significantly judged to be honest and some others dishonest: the issue of interjudge *reliability*. High agreement between observers who rated stimulus persons´ honesty on the basis of facial photographs has been reported in most studies (e.g., Berry & Brownlow, 1989; Berry & McArthur, 1985; Zebrowitz & Monteparne, 1992; Zebrowitz, Monteparne, & Lee, 1993; Zebrowitz, Voinescu, & Collins, 1996). Zebrowitz *et al.* (1996) reported that honesty-rating correlations averaged .68 for male faces and .78 for female faces.

Mueller, Thompson, and Vogel (1988) found that, unlike the dishonest faces, the hon-

est faces were also high on likability, typicality, and attractiveness. Similarly, other researchers have also found attractive faces to be perceived as more honest than unattractive faces. For example, Berry and McArthur (1985) showed that attractiveness had an impact on perceived honesty ratings. McArthur and Apatow (1983-84) concluded that physical appeal could mediate the influence of a childlike facial appearance on impressions of honesty. Berry and Brownlow (1989) reported that stimulus persons´ attractiveness was positively correlated with their perceived honesty. Berry (1990) also found that facial attractiveness predicted impressions of the honesty of photographed people. The same author again found attractiveness to influence sincerity, a construct including warmth, naiveté, straightforwardness, honesty, and kindness (Berry, 1991b). Zebrowitz and Montepare (1992) reported that, with facial maturity and smiling held constant, more attractive targets from preschoolers to older adults tended to be perceived as more honest than less attractive individuals. Zebrowitz *et al.* (1993) showed that, with a few exceptions, attractiveness of white, black, and Korean targets significantly correlated with white, black, and Korean judges´ impressions of the targets´ honesty. Zebrowitz *et al.* (1996) found that attractiveness was correlated with perceived honesty at all ages ranging from childhood through the 50s. When entered in regression analyses, attractiveness also predicted perceived honesty at various ages. These results are consistent with *the attractiveness halo effect* (e.g., Alley & Hildebrandt, 1988; Berscheid & Walster, 1974; Langlois, 1986) or *the what-is-beautiful-is-good stereotype* (Dion, Berscheid, & Walster, 1972).

Not only did Zebrowitz *et al.* (1996) find an association between honesty and attractiveness, but also between honesty and other static facial characteristics such as symmetry, eye size, and babyfacedness. This latter construct strongly correlated with perceived honesty in adulthood, puberty, the 30s and the 50s, and predicted perceived honesty in childhood, puberty, adolescence, and the 30s with all other variables controlled. This study is not the only one which has explored the role of a babyish facial appearance on perceptions of honesty. Since the mid 80s, Leslie Zebrowitz, Diane Berry and their associates have done extensive research on babyfacedness and how it influences social perceptions (see reviews by Berry & McArthur, 1986, 1988; Montepare & Zebrowitz, 1998; Zebrowitz, 1997). Impressions of honesty have usually been measured in their research. Thus, MacArthur and Apatow (1983-84) manipulated some facial features associated with babyfacedness in face composites built with a police Identi-kit. They found that increasing babyishness of faces led to linear increases in the perceived honesty of the stimulus persons. Similar results were found by Berry and McArthur (1985) using real photographs of human faces. Berry and Brownlow (1989), and Berry (1990) reported significant correlations between babyfacedness and impressions of honesty. Berry (1991b) also found facial babyishness to influence sincerity, a composite of perceived honesty, straightforwardness, warmth, naiveté, and kindness. Zebrowitz and Montepare (1992) discovered that the influence of babyfacedness upon honesty judgments extended across the life span, especially among males. Zebrowitz *et al.* (1993) found significant correlations between the facial babyishness of white, black, and Korean targets and honesty impressions of white, black, and Korean observers. In addition, research conducted by Brownlow (1993; Brownlow & Zebrowitz, 1990) indicates that babyfaced individuals are perceived as more trustworthy than maturefaced individuals. All these results are consistent with *the babyface generalization effect* (e.g., Berry & McArthur, 1986; Montepare and Zebrowitz, 1998; Zebrowitz, 1997), rooted in McArthur and Baron´s (1983) ecological theory of social perception.

In summary, observers judging facial photographs tend to agree on which stimulus persons look honest and which look dishonest. Attractive faces and babyfaced faces are more often judged to be honest than those which are unattractive or maturefaced.

Less attention has been devoted by researchers to the issue of *validity*, that is, to as-

certain whether observers´ perceptions of the honesty of target individuals are accurate. And not all the results from the few extant studies coincide in supporting or rejecting the existence of a kernel of truth in honesty judgments. On the discomfirmatory side, recent research by Andreoletti and Zebrowitz (1997, cited by Montepare & Zebrowitz, 1998) has found that young delinquents which were babyfaced accrued more criminal charges than their maturefaced fellows. In another study, middle-class babyfaced boys exhibited more negative behaviors, among which was lying, than their maturefaced peers (Zebrowitz, Collins, & Dutta, 1997; cited in Montepare & Zebrowitz, 1998). Also, Zebrowitz *et al.* (1996) failed to find significant correlations between perceived and real honesty of stimulus persons at any single age. However, they found that, among males, perceived honesty at earlier ages emerged as a positive predictor of real honesty in adulthood, and that correlations between perceived and real honesty in the later years indicated that honesty impressions tended to be accurate for men who had been stable in perceived honesty across the life span (marginally significant effect), but not for those who had been unstable in perceived honesty. However, the reverse pattern emerged for women: lower levels of real honesty at earlier ages predicted higher perceived honesty at later ages, and honesty impressions of women in their later years tended to be significantly inaccurate for women who had been stable in *real* honesty across the life span, but not for women unstable in real honesty. On the supportive side, Berry (1990) found that ratings of the honesty of a series of individuals, delivered by judges who did not know them who looked at their photographs, were related to classmates´ honesty impressions of those targets after 5 and 9 weeks of acquaintance. Also, strangers´ photography-based judgments of warmth (an aggregrate measure of warmth, kindness, and honesty) were related to males´ (but not females´) self-reported warmth and their scores in a Social Closeness Scale. From these results Berry (1990) concluded that the "data are consistent with the proposition that facial appearance may provide some accurate information about an individual´s likely behaviors" (p. 352).

Bond, Berry, and Omar (1994) challenged that conclusion. They argued that neither self- nor other-descriptions of target individuals should be used as an independent criterion of their actual honesty, because physical appearance may have such a strong impact that influences even the ratings of participants who know the person being evaluated. Or it may even be the case that people use their own physical appearance to infer their traits and attitudes, a proposition which extends Bem´s (1972) self-perception theory beyond the observation of behavior to also include the observation of one´s own appearance as a basis for making attributions about oneself. However, Berry (1990) found that attractiveness was the only facial characteristic related both to photograph-based ratings of warmth and men´s Social Closeness scores, and analyses failed to support the notion that attractiveness accounted for the relation between observers´ impressions of warmth and these scores. This does not support Bond *et al.*´s suggestion that facial appearance may be used in a similar way by both, stimulus persons and strangers, to make inferences about the stimulus persons, which would lead to an agreement between self- and strangers´ ratings. Yet we must agree with Bond *et al.*´s (1994) contention that behavioral measures should be used as the criterion against which to compare strangers´ perceptions of stimulus persons. In other words, if we are to assess whether strangers´ impressions of people´s honesty are accurate or not, we should compare these impressions with those people´s actual behavior when they have the choice to act in an honest or a dishonest way.

Two mutually-exclusive hypotheses were posed by Bond and his colleagues. According to a *social reinforcement model* perceptions of honesty would be *negatively* correlated to actual honesty. This would be so because honest-looking persons would rarely be suspected of devious behaviors, therefore they would be successful in whatever deception they at-

tempted, which would be rewarding. This would perpetuate this behavior and, in turn, repetition would increase the deceivers´ skill. The results obtained by Andreoletti and Zebrowitz (1997) and Zebrowitz, Collins, & Dutta, (1997, both cited by Montepare and Zebrowitz, 1998) described earlier lend support to the social reinforcement model.

An alternative hypothesis was based on a *self-fulfilling prophecy model*, which predicted that perceptions of honesty would be *positively* correlated with actual honesty. This would be so because, according to this view, honest individuals would be treated in such a way that they would become honest, and dishonest individuals would be treated in a way that would made them become dishonest (for reviews on the self-fulfilling prophecy see Cooper & Good, 1983; Darley & Fazio, 1980; Miller & Turnbull, 1986; Snyder, 1984). The aforementioned results obtained by Zebrowitz *et al.* (1996), showing that males´ earlier honesty predicted later real honesty, and that among males the relation between perceived and real honesty at later ages was significant among those who had been perceived as honest across their life span, are consistent with this behavioral confirmation effect.

To test their hypotheses Bond *et al.* (1994) took photographs of 133 undergraduate students. Later, these students were provided with written descriptions of 8 experiments, 6 of which involved deception, and had to indicate whether they would be willing to participate in each of them or not. After that, participants were provided with an opportunity to write a deceptive letter addressed to an unknown student. Then the researchers combined the seven deceptive items (six experiments plus the deceptive note) to form a $0 - 7$ scale of willingness to perform deceptive behaviors. Later on, participants´ pictures were shown to a sample of 22 observers unacquainted with them, who judged in a scale the extent to which each person in the photographs looked honest or dishonest. A separate group of individuals rated the faces in terms of their attractiveness and babyfacedness. Results supported the self-fulfilling prophecy model: honest-looking individuals were less likely to volunteer to participate in deceptive experiments than their dishonest-looking fellows. Somewhat surprisingly neither physical attractiveness nor facial babyishness accounted for these results.

This kind of research may have important implications in other areas of inquiry. For instance, both Bond *et al.* (1994) and Zebrowitz *et al.* (1996) explicitly mention the relevance of their work for the prolific research area of the detection of deception from nonverbal cues (for recent reviews on the topic see, e.g., Ekman, 1992; Miller and Stiff, 1993; Vrij, 1998, 2000). In 1979, Zuckerman, DeFrank, Hall, Larrace, and Rosenthal found what they termed a *demeanor bias* in their senders: some where consistently judged as honest and some as deceptive, regardless of whether they lied or told the truth. The existence of a demeanor bias has been confirmed by later research conducted by Bond, Kahler, and Paolicelli (1985). As it is conceptualised by Zuckerman *et al.* (1979), that bias would depend on some internal characteristics influencing the sender´s perceptible *demeanor* which, in turn, would determine observer´s ratings. Indeed, some authors have tried to see the influence of some personality traits and social skills of the sender upon the observers´ credibility judgments (e.g., Geis & Moon, 1981; Miller, deTurck, & Kalbfleisch, 1983; Riggio & Friedman, 1983; Riggio, Tucker, & Widaman, 1987; Riggio, Tucker, & Throckmorton, 1987; Vrij, 1992; Vrij & Winkel, 1993), assuming that these traits influence in some way the behavior displayed by the communicator (for empirical tests of this assumption see Riggio, Tucker, & Widaman, 1987; Vrij, Akehurst, & Morris, 1997). However, as suggested by Bond *et al.* (1994), Bond and Robinson (1988), and Zebrowitz *et al.*, (1996), it may be the case that these biases originate in static facial characteristics that would bestow on the individual "an innocent- or guilty-looking visage" (Bond & Robinson, 1988, p. 304). In this case, the biased judgments of credibility would depend directly upon the sender´s appearance, instead of de-

pending on some personality traits or social skills which would influence behavior. This hypothesis was recently tested by Masip, Garrido, and Herrero (1999, Study 2), who failed to find support for it, probably because, as they themselves acknowledged, the faces they used were not chosen on the basis of differences in attractiveness, facial maturity, or perceived age.

This paper is a replication of the study conducted by Bond *et al.* (1994). Replications are necessary before drawing strong conclusions from experimental results. In addition, cross-cultural confirmation of findings obtained with North American samples is needed before such findings be generalised to people from other countries. In this regard, we used Spanish participants both as stimulus persons and as judges who had to rate targets´ faces in a series of dimensions. Finally, our procedure does not coincide exactly with that of Bond *et al.*, as we introduced some minor modifications in order to improve upon the original procedure. A first modification consisted in the following: as in the original study, respondents were asked whether they would participate in the experiments or not; but, in addition, those who said they would were required to indicate on a scale the strength of their willingness to participate. This would provide us with a more fine-grained measure of participants´ readiness to perform the deceptive experimental tasks described in the forms. Second, in order to take into account participants´ willingness to collaborate in the nondeceptive experiments in addition to their readiness to collaborate in the deceptive ones, a new dependent measure was considered which consisted of the combination of the seven deceptive items (range 0 – 7) *minus* the combination of the two truthful items (range 0 – 2; final range: -2 – 7). (See method section).

Third, for the same reason that respondents´ willingness to participate in the nondeceptive experiments was substracted from their willingness to take part in the deceptive ones, ratings of the strength to do so had to be substracted as well. In this way a new dependent variable was created.

Fourth: In order to prevent participants´ compliance or noncompliance with the requirement to write the deceptive letter due either to normative (e.g., Asch, 1951; Deutsch & Gerard, 1955; Moulton, Blake, & Olmstead, 1956) or informational (e.g., Sheriff, 1936; Rohrer, Baron, Hoffman, & Swander, 1954) social influence, an additional requirement was added for those participants who did not want to write the letter –they were required to write the reasons for their decision. That is, since we planned to collect the data in a classroom with a large number of students filling in the form at the same time, maybe some of them would write (or would not write) the letter because they would see other students doing so (or not doing so). By asking respondents either to write the letter or express the reasons why they chose not to, we would have all participants writing something. Any of them could see his or her companion writing, but he or she would be unable to know whether it was the letter or the reasoning his or her refusal to word it. An additional, although crucial, benefit would be that we could examine the reasons given by noncompliant participants: Did they refuse to write the letter, as Bond *et al.* assumed, for ethical reasons? Or were there reasons other than ethical concerns for them not to write it?

Fifth, in addition to the honest / dishonest scale judges in the study done by Bond *et al.* completed, a deceptive / truthful scale was completed by our respondents[1].

[1] This scale was included because the meaning of the English word *honesty* is not exactly the same as that of its Spanish equivalent. According to The Concise Oxford Dictionary, *honesty* is "1. being honest. 2. truthfulness.", and *honest* means "1. fair and just in character or behaviour, not cheating or stealing." but also "...2. free of deceit and untruthfulness, sincere." Thus, the English word *honesty* is very close to the words *truthfulness*, *sincerity* or *veracity*. This is not so for the Spanish word for *honesty*, namely "*honestidad*". The Dictionary of the Real Academia Española de la Lengua defines it as "quality of being honest", and defines *honest* ("*honesto*") as "1. Decent or decorous. 2. Modest or shameful. 3. Reasonable, fair. 4. Upright." Thus, the Spanish word for *honesty* is farther from veracity or truthfulness than its English counterpart, being closer to a general quality of character indicating uprightness. To properly measure Spanish observers´ impressions of participants´ truthfulness, a

Finally, two additional scales were included in the questionnaire where participants had to indicate their willingness to participate in the experimental procedures: they were asked to indicate in a 1 (truthful) to 7 (deceptive) scale: (a) to what extent people who knew them well (i.e., their family or their flatmates) thought they were truthful or deceptive, and (b) to what extent they actually were truthful or deceptive individuals. This would permit us to examine a series of questions. First: if, as could be expected in view of Bond *et al.*´s results, we found support for a self-fulfilling prophecy model, this conclusion would be strengthened if in addition participants´ perceptions of others´ views of their honesty were correlated with their self-reports. Second: as said above, some researchers (e.g., Berry, 1990) used self-reports or acquainted others´ opinions as criteria against which to compare observers´ photograph-based impressions. However, the question remains as to how valid these criteria are. That is, to what extent individuals who say they are honest or dishonest are actually so? Similarly, to what extent are actually honest or dishonest those individuals whose friends or relatives say they are so? We could assess this by calculating the correlations between participants´ self-ratings of honesty and close acquaintances´ views of their honesty (as reported by the participants) on the one hand, and participants´ actual honesty (i.e., their willingness to participate in the deceptive studies) on the other. Third, are Bond *et al.* (1994) right when they contend that close acquaintances´ impressions as well as participants´ self-evaluations may be influenced by targets´ facial appearance? This could be tested by examining the correlations between participants´ self-ratings of honesty and close acquaintances´ views of their honesty (as reported by the participants) on the one hand, and observers´ photograph-based impressions on the other.

In summary, the aims of the present study were the following: first, to measure observers´ agreement in their photograph-based impressions of others as honest or dishonest (the reliability question); second, to test whether these impressions are accurate or not (the validity question) –a positive correlation between perceived and real honesty would support the self-fulfilling prophecy model, a negative correlation would support the social reinforcement model, neither model would be supported if no correlations between real and perceived honesty emerged; third, to explore the role of attractiveness and facial babyishness on perceived and real honesty, as well as on their relation; fourth, to examine the accuracy of participants´ self-reports of their honesty as well as close acquaintances´ views of targets´ honesty (as reported by the participants); fifth, to explore whether acquaintances´ and participants´ self-reports are based upon participants´ facial appearance or not. In addition, correlations among facial maturity and attractiveness were calculated in order to investigate whether they were independent or not. Also, our data permitted us to assess to what extent undergraduates have reservations about lying.

Method

Overview

Participants (undergraduate students) were asked to complete a questionnaire where eight experiments were described. Six of them required that participants engaged in deceptive behaviors. They had to indicate what experiments they would be willing to participate in as research assistants. In addition, they were given the opportunity to write a deceptive note. These procedures enabled us to measure respondents´ willingness to engage in deceptive behaviors. Later on, participants gave us their module cards, which included a facial photograph of them. These pictures were scanned and shown with a PowerPoint presentation to two samples of observers who were unacquainted with the stimulus persons. The first sample of observers rated participants´ honest- or dishonest-looking appearance, as well as

ers´ impressions of participants´ truthfulness, a truthfulness / deceptiveness scale had to be used.

whether they looked truthful or deceptive. The second sample rated participants´ facial attractiveness and facial maturity.

Participants

Participants were psychology students at a Spanish university who were taking the module "Psychology and Law". As we shall see later, they were required to complete a questionnaire as well as to hand over their module cards, which included personal information and a photograph of the card holder, to the module lecturer.

The final number of questionnaires completed by the students was 105, of which 15 were useless because the students who had filled them in had not handed over their module cards (hence, no photograph of them was available). One further questionnaire was rejected because the corresponding photograph was a useless black-and-white photocopy of the original picture. Thus, 89 questionnaires were left.

The number of students who gave us their module cards was 97. The black-and-white photocopy mentioned above was excluded, as were seven further pictures of participants who had not handed over the completed questionnaires. Of these seven pictures, four were later included in the PowerPoint presentation as practice items. However, the useful pictures in that presentation were 89. Later, four of them were excluded from analyses, since the corresponding participants had left unchecked some items from the questionnaire.

Thus, the final number of participants was 85 (80 females and five males[2]; age range: 20 – 26 years, $M = 21.46$).

[2] No gender differences were found by Bond *et al.* (1994), thus no effort was made to select a sample with the same amount of males and females in it. In any case, caution is warranted before generalizing our results to male populations.

Procedure

Actual honesty and facial photographs. On arriving at the lecture room at the beginning of the academic term, those undergraduate students of psychology who were taking the module "Psychology and Law" were told they had to participate as research assistants in two of our deception detection experiments. A questionnaire with written descriptions of 8 studies was given to them. They were required to indicate whether they were willing to collaborate in each study or not by checking the yes or the no square which appeared after each experimental procedure description in the questionnaire. In addition, those who indicated they agreed to take part in a specific study had to indicate on a scale ranging from 1 to 10 the extent to which they were willing to participate. Six of the experiments in the questionnaire involved deception, whereas two of them did not. Descriptions of the experimental procedures were taken from Bond *et al.*´s (1994) paper. Those involving deception required from participants to falsely promise a reward to another person, to feign pain, to give another student false feedback about his or her results in a personality test, to simulate suffering from a mental illness, to concoct answers to impossible questions, or to falsely tell a student suffering from speech anxiety that he or she would have to give a public speech. Procedures not involving deception required students to make personality judgments based upon nonverbal behaviors or to deliver experimental instructions in a memory experiment. Participants were asked to agree to participate in at least two experiments –otherwise we would have run the risk that most students refused to participate in any study, either deceptive or not. However, they were encouraged to agree to as many studies as possible to facilitate scheduling. They were notified in advance that most experiments required deceiving someone else but that, in case they had reservations about lying, some of the studies did not involve deceptive practices.

Using a procedure similar to Bond *et al.*´s (1994), we provided participants with an addi-

tional opportunity to engage in deceptive behavior. They were falsely told that the university had started up a new "matchmaking" service. With this aim in view, some students had purportedly described themselves while being video recorded so that, later on, these video tapes could be shown to other students who were looking for someone to date. We asked our participants to act as if they had watched the original tapes and to write a deceptive letter saying they wanted to have a date with a person in one of these tapes. This letter would ostensibly be given to a user of the video dating service, and whether he or she believed its content would be assessed. Participants in our study were told we were interested in studying whether they were able to deceive the clients of this new university dating service. It was stressed that they had no obligation to write the deceptive letter, but if they chose not to they had to write down their reasons for not wanting to participate. All participants were provided a sheet of paper to write either the deceptive note or their reasons for not writing it.

Finally, participants had to indicate on two 7-point scales which were included in the questionnaires to what extent close acquaintances thought they were truthful (1) or deceptive (7), and how truthful (1) or deceptive (7) they were in reality. Students were told this was an important control variable to be taken into account when considering their potential ability to deceive participants in the deceptive experimental procedures.

The whole data collection session was carried out twice, since students of two classes participated (allocation of students to one or the other class is based upon an alphabetic criterion). After all students had finished their tasks the questionnaires were collected and participants were thanked. A large group of students who did not attend the lectures where data were collected asked us individually to give them the forms so that they could complete them at home. We gave them blank forms they completed before returning them. The number of students in class A was 37, 16 attended class B, and 32 filled in the forms at home. One-way analyses of variance were conducted to check whether students´ willingness to perform deceptive behaviors varied as a function of their group (A, B, or home). No significant effects emerged, therefore this variable was not taken into account in subsequent analyses[3].

At the beginning of the semester students at our university must hand in the module cards to their lecturers. These are cards where personal information on the student (gender, date of birth, postal address, telephone number, etc.) is provided, along with a passport-type color photograph of the card holder. Each lecturer must receive a card from each student taking his or her module. Lecturers may use the back side of the cards to write notes about the students´ performance, lecture attendance, or whatever. Receiving the cards from those who were taking our module on psychology and law would allow us to easily obtain facial photographs of our experimental participants.

After all students had given their forms back and the module cards had been collected, the participants were debriefed during a lecture on social perception of faces and its implications for legal procedures. In that lecture, the background of the study was described in detail. In addition, participants were asked to give us their permission to show their photographs to other students that would rate their honesty,

[3] As we shall see later, four measures of participants´ real honesty were used in this study: (a) number of deceptive experiments participants were willing to participate in, (b) this amount minus the number of truthful experiments participants agreed to take part in, (c) a quantitative measure of participants´ willingness to participate in the deceptive studies, and (d) this measure minus students´ willingness to collaborate in the nondeceptive procedures. Respondents´ group (class A, class B, home) was introduced in four univariate analyses of variance which were performed on the four actual honesty measures. Results were, respectively, F (2, 82) = .41, p = .667; F (2, 82) = .10, p = .902; F (2, 82) = .39, p = .680; F (2, 82) = .10, p = .909. Two similar analyses were performed on students´ notions of others´ views of their truthfulness, and their own truthfulness ratings. Results indicated that the participants´ group had no influence on these two variables either; respectively: F (2, 76) = 1.01, p = .370; F (2, 76) = .49, p = .616.

truthfulness, attractiveness, and facial maturity. None of the participants refused.

Observers´ impressions of targets´ faces. All the photographs were scanned by a research assistant and included in a PowerPoint presentation. Eighty-nine stimulus-person photographs plus four practice pictures were in the presentation. Each photograph was shown for 6 seconds (Bond *et al.*, 1994), and observers had 6 further seconds between one photograph and the next to check the scales by which pictures were to be evaluated. A sound (Microsoft´s "Clic.wav") could be heard as a photograph appeared on screen, and a different one (Microsoft´s "Sports-Bip.wav") as it disappeared. The sounds were included so that if participants were not looking at the screen, but were completing the scales, they would know the next slide was being projected. All pictures had a size of 142 X 226 pixels and appeared in the center of the screen. The rest of the screen was dark. Each picture was accompained by a written verbal label: "Photo number X", which appeared above it. After each photograph was shown, a black screen with the message "Please assess this photograph" was visible during the 6 seconds participants had to check the scales for each picture.

Twenty-four undergraduate students of criminology at the faculty of law (three did not report their age or gender, 18 of the remaining students were females and three were males[4]; age range: 18 – 26, $M = 21.14$) were shown the presentation and had to rate the truthfulness and honesty of the persons in the photographs. For this purpose, they received a questionnaire with two bipolar scales, ranging from 1 to 7, for each photograph. Endpoints were labelled "truthful" (1) and "deceptive" (7) in one scale, and "honest" (1) and "dishonest" (7) in the other. After looking at each photograph for six seconds observers had six further seconds to assess the stimulus person by checking the appropriate rating on the scale. Observers were informed of the procedure, and then they were shown four practice items which they had to evaluate. Then the presentation was stopped and participants were asked whether they had had any problem in rating the pictures, and whether they could clearly see the slides and hear the sounds. Also, they were invited to ask any question regarding the procedure. No questions were asked, therefore the experimental session began. A countdown sequence of slides showing numbers from 5 to 1 at a rate of one per second was shown before the first experimental photograph was projected. This was done to ensure observers would be ready when the first photograph appeared. The presentation was projected on a screen in front of the students in a classroom at the faculty of law. After the presentation was over, students were thanked and debriefed while being given a lecture on social perception of faces and its legal implications.

Exactly the same procedure was used with another group of 18 students, taken from the same population and with similar age and gender composition. However, this second sample did not judge stimulus persons´ honesty and truthfulness, but their attractiveness and babyfacedness. In order for them to be capable of judging facial maturity, prior to the experimental session they attended a 20-minute lecture about the physical features of babyish faces and those of mature faces. No mention was made during this lecture of the attributional and social consequences of having a babyish or a mature face –this information was provided during the debriefing session. After that initial lecture they completed the experimental task.

It was supposed observers from both groups were unacquainted with the stimulus

[4] An equal number of males and females with similar background characteristics was not available at the time data were collected. Although similar number of male and female observers have been used in extant research on babyfacedness and honesty impressions, in general researchers have not reported whether raters´ gender differences had any effect on the dependent measures. Berry (1991b) did introduce sex of judges as one of the factors in an analysis of variance examining perceived sincerity. She did not report any main effect or interaction where that variable was involved. In any case, we think caution is warranted before generalizing our results to male observers´ populations.

persons. However, they were instructed that in case they recognized someone they were to leave the scales where that person had to be evaluated unchecked. For this reason, or due to any other reason, one respondent did not rate the attractiveness and babyfacedness of four stimulus-persons, and another left one attractiveness scale and two babyfacedness scales unchecked. These empty cells were filled in with the mean value of the other respondents´ ratings of the specific faces. The same strategy was used to complete two missing ratings of the veracity scales and two others of the honesty scales.

Results

After all data had been collected we noticed that four respondents had left some scales unchecked where they should have indicated whether they were willing to participate or not in the experiments which were described. These participants were excluded from the sample, and all analyses were performed on the remaining 85. (See the "participants" section).

Descriptive analyses

Participants. On average, participants were willing to participate in 4.40 (SD = 1.46) deceptive experiments (range: 0 – 7, i.e., 6 deceptive experiments plus the deceptive letter). This figure is somewhat higher than Bond *et al.*´s (1994) 3.95. The difference may be due to the fact that, unlike Bond and his colleagues, we required our participants to agree to participate in at least two experiments, either deceptive or nondeceptive. Most respondents agreed to participate in four (N = 25, 29.40 % of participants) or five (N = 25, 29.40 % of participants) deceptive procedures.

The number of truthful experiments students were willing to participate in was substracted from the number of deceptive experiments they agreed to collaborate in. We then had a scale ranging from –2 to 7. The average rating of participants in this scale was M = 2.67, SD = 1.56. Most observers had a rating of 3 (N = 31, 36.5 % of participants). None had a value of 7.

Table 1 shows the number of participants who agreed or disagreed to take part in the studies. It is apparent that most students agreed to participate in most experimental procedures, either truthful or deceptive. Exceptions were simulating pain, where most students disagreed, $\chi^2_{(1)} = 23.82$, $p = .000$, and feigning suffering from a mental illness, where no significant differences emerged between the number of students willing to participate and those who did not want to participate $\chi^2_{(1)} = 2.65$, $p = .104$. The right column of Table 1 includes the proportion of participants who agreed to collaborate in each procedure. A one-way ANOVA on these results, where experiment was included as the factor, showed some studies were chosen more often than others, $F(8, 764) = 24.70$, $p = .000$. As we expected, the most frequently chosen experiment to participate in was a nondeceptive one, namely, the memory experiment. No significant differences were found between participants´ willingness to collaborate in this study and their willingness to participate in the other nondeceptive procedure (social judgment experiment), as shown by the post-hoc Scheffé test (p = .990). Also, it is somewhat surprising that students´ tendency to agree to participate in a number of deceptive experiments (false feedback, bogus reward, rigged contest, and public speech) did not differ significantly from their tendency to agree to take part in the nondeceptive procedures.

As stated above, in addition to asking participants to indicate whether they would be willing to participate in each experimental procedure or not, we also asked those who said they would to indicate *the extent* to which they would agree to participate, on a continuous scale ranging from 1 (not very much) to 10 (very much). When introducing the data in the computer 0 values were introduced in this variable for those who indicated they would *not* participate in a given study. Participants´ ratings for the deceptive experiments were added

up, and the final value was divided by 10. We then had a scale ranging from 0 to 6 (not 7, since no scale was provided for the deceptive letter experiment) where 0 meant the respondent was not willing at all to participate in the deceptive experiment and 6 meant he or she was completely willing to do so. Participants´ mean rating in this scale was M = 2.51, SD = 1.10, range: 0.00 – 5.10. This seems to indicate that, although participants were willing to participate in much of the deceptive procedures (4.40 out of 7, as said before), their desire to do so was not particularly strong (2.51 in a scale from 0 to 6).

Table 1: Frequencies of participants who agreed or disagreed to participate in each experiment, chi-square analyses, and proportion of those who agreed.

	Number of participants (percentage)		Chi-square analyses		Proportion of respondents who agreed*
	Agreed	Disagreed	Chi-square value	p	
Memory (no deception)	77 (90.6)	8 (9.4)	56.01	.000	.91[a]
False feedback	72 (84.7)	13 (15.3)	40.95	.000	.85[ab]
Bogus reward	71 (83.5)	14 (16.5)	38.22	.000	.84[ab]
Social judgment (no deception)	70 (82.4)	15 (17.6)	35.59	.000	.82[ab]
Rigged contest	66 (77.6)	19 (22.4)	25.99	.000	.78[ab]
Public speech	56 (65.9)	29 (31.4)	8.58	.003	.66[abc]
Letter	54 (63.5)	31 (36.5)	6.22	.013	.64 [bc]
Mental illness	35 (41.2)	50 (58.8)	2.65	ns	.41 [cd]
Pain simulation	20 (23.5)	65 (76.5)	23.82	.000	.24 [d]

* Only means with a different superscript are different at $p < .05$. Post-hoc Scheffé tests were used to check which values were significantly different from each other.

In order to also take into account respondents´ strength of their willingness to participate in the nondeceptive experiments, their ratings in the scales of the nondeceptive studies were added up, the final value was divided by 10, and the resulting value was subtracted from respondents´ desire to participate in the deceptive procedures (see previous paragraph). The resulting scale could theoretically range from –2 to 6. Participants´ mean value in this scale was M = 1.17, SD = 1.28, range: -2.00 – 5.00.

In fact, all four measures (i.e., number of deceptive experiments participants agreed to participate in, number of deceptive experiments participants agreed to participate in minus the number of truthful studies they agreed to participate in, respondents´ degree of willingness to participate in the deceptive experiments, and their degree of willingness to participate in the deceptive experiments minus their degree of willingness to participate in the truthful) were highly correlated (Pearson correlations ranged between .76 and .95, all ps = .000).

Observers. Interjudge reliabilities were calculated for observers´ impressions. Alphas were .78 for truthfulness, .76 for honesty, .91 for attractiveness, and .87 for babyfacedness. Since interjudge agreement for these four variables was high, the mean rating each stimulus person received on each of them was calculated.

Attractiveness and facial babyishness showed a marginally significant correlation, r = .19, p = .082. This may be due to the fact that almost all the photographs depicted women, and women are in general more babyfaced than men. Since prototypical faces are normally

judged as more attractive than the non-prototypical (e.g., Langlois & Roggman, 1990), it is reasonable that babyfaced female faces be considered as more attractive than less babyish female faces.

A highly significant positive correlation was found between perceived honesty and truthfulness, $r = .95$, $p = .000$. This indicates that individuals whose character in general is perceived negatively (see footnote 1) are also though to be deceptive, while those who are considered to be upright are also regarded as truthful persons. No other correlation was significant: the relation between attractiveness and perceived truthfulness was $r = .09$, $p = .431$; the correlation between attractiveness and honesty impressions was $r = .04$, $p = .709$; the relationship between facial babyishness and impressions of truthfulness was $r = .10$, $p = .384$; and facial babyishness correlation with perceived honesty was $r = .08$, $p = .463$.

Main Analyses

Accuracy of observers´ impressions. In order to test whether honesty impressions based on facial photographs were accurate or not, correlations between observers´ ratings of stimulus persons´ honesty and the actual honesty of those stimulus persons were calculated. As stated above, we had four measures of real honesty: (a) number of deceptive experiments participants were willing to participate in (this was Bond *et al.*´s main variable), (b) this number minus the number of truthful experiments participants agreed to take part in, (c) a quantitative measure of participants´ willingness to participate in the deceptive studies, and (d) this measure minus students´ willingness to collaborate in the nondeceptive procedures. Although, as we have said, all four variables were strongly correlated, separate analyses were conducted for each. The correlation of honesty impressions with the first real honesty measure was $r = .06$, $p = .603$, its correlation with the second measure was $r = .08$, $p = .460$, with the third measure $r = .01$, $p = .965$, and with the fourth actual honesty measure $r = .06$, $p = .614$. In summary: it seems observers´ impressions of targets´ honesty had little to do with their real honesty. Results do not support either the self-fulfilling prophecy model nor the social reinforcement model.

Participants were asked whether they would lie or not. But honesty, as explained above (see footnote 1), is for Spaniards something more vague and general than the tendency not to lie or cheat. Although , as described earlier, observers´ ratings of target individuals´ honesty and truthfulness were strongly correlated, a more proper test of our prediction of a consistent relationship between observers´ impressions and students´ willingnesss to lie should involve calculating the correlations between observers´ *truthfulness* impressions of participants and the real honesty of these participants (i.e., their willingness to *lie* in the context of collaborating in experimental procedures). These analyses were performed. The correlation of observers´ impressions of target participants´ truthfulness with our first real honesty measure was $r = .07$, $p = .514$, its correlation with the second measure was $r = .10$, $p = .354$, and its correlation with the third measure was $r = .03$, $p = .781$, while with the fourth measure the correlation was $r = .08$, $p = .449$.

In summary: Observers´ impressions of the honesty of target participants´ honesty and truthfulness based on their facial photographs was unrelated to those participants´ willingness to engage in deceptive practices.

Participants´ self-ratings of truthfulness. Participants were asked to indicate, according to two scales ranging from 1 (truthful) to 7 (deceptive), to what extent close acquaintances thought they were truthful or deceptive, and to what extent they were actually truthful or deceptive. These two measures were strongly correlated, $r = .60$, $p = .000$, showing that people´s self-perceptions of their truthfulness is in line with how they think others view them. Had we found support for a self-fulfilling prophecy model, this result would have strengthened that support.

We also measured whether participants´ notions of their own truthfulness were accurate. In order to do so, correlations were computed between participants´ self-ratings of truthfulness and the four measures of their real honesty. The correlation with the first actual honesty measure was $r = .19$, $p = .09$, with the second measure $r = .17$, $p = .142$, with the third $r = .04$, $p = .702$, and with the fourth $r = .07$, $p = .525$. That is: observers´ self-views about their truthfulness do not coincide with their real tendency to lie, especially when corrections for their tendency to tell the truth are introduced, and fine-grained quantitative scales of their willingness to deceive are used. This questions researchers´ use of self-reports as an independent measure of participants´ real honesty against which to compare observers´ impressions. Instead, others´ perceptions seem to be more useful: correlations of close acquaintances´ views of participants –according to participants´ reports of these views– with students´ actual honesty were calculated. All four correlations were statistically significant: The relationship of others´ views about stimulus persons´ truthfulness with the first actual truthfulness measure was $r = .28$, $p = .01$, its relation with the second measure was $r = .31$, $p = .005$, with the third measure it was $r = .22$, $p = .049$, and with the fourth $r = .26$, $p = .022$. Thus, close acquaintances´ views of the participants´ truthfulness, as perceived by the participants themselves, are fairly accurate.

Bond *et al.* (1994) suggested that both close acquaintances´ ratings of stimulus persons´ honesty and these stimulus persons´ self-ratings of honesty could be influenced by targets´ facial appearance. In order to test this, a series of correlations were calculated. The first was that between participants´ self ratings of truthfulness and photograph-based impressions of honesty. This correlation was not significant: $r = .15$, $p = .188$. Second, the correlation between perticipants´ self ratings of truthfulness and photograph-based impressions of truthfulness was calculated. This correlation was not significant either: $r = .12$, $p = .288$. Thus, it seems people do not base their self-views about how honest or dishonest they are on their honest- or dishonest-looking facial appearance. Does their facial attractiveness or babyishness have an influence on participants´ self-view as honest or deceptive persons? Since no significant correlations were found between attractiveness and babyfacedness on the one side and perceived honesty and perceived truthfulness on the other, no significant correlations bertween self-perceived truthfulnes and self-perceived honesty were expected. And, in fact, they were not found: correlations were $r = –.02$, $p = .877$ for the association between self-reported truthfulness and attractiveness, and $r = –.06$, $p = .614$ for the association between self-reported truthfulness and babyfacedness.

Next, the correlation between participants´ reports on the extent to which close acquaintances viewed them as being honest or deceptive and photograph-based impressions of their honesty and truthfulness were calculated. Values were, respectively, $r = .15$, $p = .198$, and $r = .16$, $p = .157$. Thus, it seems stimulus persons´ close acquaintances do not base their views about how honest or dishonest targets are on targets´ honest- or dishonest-looking facial appearance. Does stimulus people´s facial attractiveness or babyishness have an influence on acquaintances´ perceptions of their honesty or deceptiveness? It does not: the correlation between acquaintances´ ratings of targets´ truthfulness and their facial attractiveness was $r = .10$, $p = .361$; the correlation between acquaintances´ ratings of stimulus persons´ truthfulness and their facial babyishness was $r = –.034$, $p = .767$. Probably, as acquaintances´ ratings of stimulus persons´ honesty are accurate and do not correlate with photograph-based impressions of their honesty, truthfulness, attractiveness or babyfacedness, they do not base their impressions on targets´ facial appearance, but on their actual behavior. In line with this, correlations between attractiveness and actual honesty were in general non-significant (for the four honesty measures, in order: $r = .16$, $p = .151$; $r = .22$, $p = .046$; $r = .09$, $p = .437$; $r = .15$, $p = .161$), as were correlations between babyfacedness and actual hon-

esty (r = .07, p = .514; r = .07, p = .556; r = –.03, p = .822; r = –.03, p = .815).

Additional analyses

In addition to indicate whether they would participate in a series of deceptive experiments, participants were given the oportunity of engaging in a deceptive behavior at that very moment. This behavior was writing a deceptive note to an unknown person asking him or her for a date. It was assumed that participants´ rejection to write the deceptive note would be based on ethical reservations about lying. However, as we asked participants who were not willing to write the letter to express instead their reasons for not doing so, we could examine whether our assumptions about their ethical concerns were right or not.

Out of 85 respondents 54 wrote the letter and 31 did not. Among these, three gave no reasons for not writing the deceptive note, 19 gave one reason, and nine gave two reasons. The most frequently mentioned reason for not writing the letter had little to do with ethical concerns: it was participants´ perceptions of their lying ability or, in other words, their perceived self-efficacy for lying, which was mentioned by 22 participants. Only 11 respondents argued they did not write the letter because they had ethical reservations about lying. Four respondents also gave reasons other than low perceptions of self-efficacy or ethical concerns. These results question Bond *et al.*´s, as well as our own implicit assumption that respondents who did not write the letter did not do so because they thought lying was wrong or immoral. And this indicates that, perhaps, the actual-honesty measure used by Bond *et al.* (1994) (number of deceptive experiments participants were willing to participate in, including the letter) as well as those honesty measures used in the present study which include the letter (i.e., number of deceptive experiments participants were willing to participate in, including the letter, and this very measure minus the number of truthful experiments participants agreed to participate in), are not valid enough, because participants may be dishonest and yet may not write the letter because they feel they are not good at lying. For this reason, the most valid measures of actual honesty of the present study are those which do not involve the writing-letter data, that is, the measures of participants´ degree of willingness to engage in deceptive behaviors as expressed in the quantitative scales, or these measures minus participants degree of willingness to participate in the honest experiments. These two measures have the additional advantage of being more fine-grained than the others, since here respondents did not give merely dichotomic responses but expressed their willingness to participate in 10-point quantitative scales. However, as described above, results were very similar regardless of the real honesty measure which was taken. This is why we have included results obtained with all four in the present report, instead of only the results obtained with the two which seem to be most valid.

Discussion

Although, in line with previous research, we have found observers tend to agree in their perceptions of others as honest or dishonest, we found that these perceptions were inaccurate. Observers´ impressions of participants´ honesty as evaluated on an honest / dishonest scale as well as on a truthful / deceptive scale were unrelated to participants´ willingness to perform deceptive behaviors, regardless of whether such willingness was measured by considering the total number of deceptive studies participants accepted to participate in (substracting or not the amount of nondeceptive experiments participants agreed to take part in) or by taking into account the strength of the participants´ willingness to participate in the experiments, as measured on a 10-point continuous scale (substracting or not the strength of their willingness to collaborate in the nondeceptive studies). No support was found neither for the self-fulfilling prophecy model nor for the social reinforcement model: there was no relation between facial appearance and willingness to deceive. As described in the

ingness to deceive. As described in the introduction to the present paper, previous research results are contradictory in their answers to Berry and Brownlow´s (1989) question as to whether physiognomists were right. However, since the most essential aspects of Bond *et al.*´s (1994) experimental procedure were closely mirrored in our study it is surprising that our results do not confirm theirs. An explanation which could account for this discrepancy concerns cultural differences: maybe facial stereotypes are accurate among North American college students and inaccurate among Southern European undergraduates. This could be based on social practices which would put in motion a self-fulfilling-prophecy process among Americans, but not among Spaniards, but research is needed to test this tentative explanation. In any case, these considerations stress the importance of doing cross-cultural research both in the nonverbal detection of deceit (e.g., Bond, Omar, Mahmoud, & Bonser, 1990; Bond, Omar, Pitre, Lashley, Skaggs, & Kirk, 1992; Cody, Lee, & Chao, 1989; see also Ekman, 1997, and Znakov, 1997) and in the social perception of faces (e.g., Keating, Mazur, & Segall, 1981; Zebrowitz, Montepare, & Lee, 1993). In any case, we think strong conclusions should not be drawn either from our results nor from Bond *et al.*´s until more research on the accuracy of honesty impressions based on facial photographs has been conducted both in the USA and in European countries.

Not only were correlations between honesty impressions and real honesty nonsignificant, but the same was true for correlations between facial babyishness or attractiveness and real honesty as well. This means that psychosocial processess analogous to those predicted to play a role in the relation between photograph-based honesty impressions and actual honesty (i.e., social reinforcement processes or self-fulfilling prophecy processes) are not in operation concerning babyfacedness or attractiveness either. For example, in the same way it was suggested honest-looking individuals would not be suspected of devious behaviors, therefore they would succeed when deceiving, which would in turn be rewarding making them deceive again and again (the social reinforcement model), it could also have been suggested that, similarly, attractive individuals, on the basis of the attractiveness halo effect, would rarely be suspected of devious acts, thus suceeding at deceiving others, which would be gratifying making them behave in a similar way again and again. Our data, however, refute any possible hypothesis concerning a link between attractiveness or babyfacedness and actual honesty.

Also apparently discrepant with the bulk of previous research was our finding of no relationships between neither attractiveness nor babyfacedness and photograph-based honesty impressions. Despite the fact that, as explained in the introduction, most studies have found that targets´ facial babyishness and attractiveness have an influence upon how honest they are considered to be by unacquainted observers, some exceptions have also been found. For example, Zebrowitz and Montepare (1992) found that, for female targets, correlations between babyfacedness and honesty impressions were not significant for 5th graders, 8th graders, and young adults. Also among women, attractiveness was unrelated to honesty impressions at preschool ages, 5th grade, 8th grade, and young adulthood. If we consider that almost all of our stimulus persons were young adult females (there were only five men in the sample), our results are in line with those of Zebrowitz and Montepare (1992).

Two interesting questions examined in the present investigation are how accurate targets´ self-evaluations of honesty are, as well as how accurate acquainted others´ assessments are. These questions are important because in some previous studies (e.g., Berry, 1990) self-ratings and close acquaintances´ impressions were used as the independent criteria against which to compare observers´ photograph-based ratings. Our results showed that self-reports on one´s own honesty are not valid measures of participants´ actual honesty, understanding actual honesty as the willingness not to engage in deceptive behaviors. Close acquaintances´

views, as reported by the participants, are indeed more valid, since they significantly correlated with our four measures of real honesty. Probably acquaintances´ perceptions are right because they are based on the observation of participants´ actual behavior in their daily life, not on the participants´ facial appearance. Indeed, when acquaintances´ views of targets´ truthfulness were correlated with photograph-based honesty impressions no significant associations were revealed. Similarly, observers´ self-reported truthfulness was also independent from unacquainted observers´ honesty impressions. This indicates that facial stereotypes are not strong enough to influence either self-perceptions or close acquaintances´. However, we must admit that participants´ reports of close acquaintances´ views may be inaccurate. That is, participants may be unaware of how truthful or deceptive their friends and family think they are. This is unlikely, since in that case no explanation would be available to account for the significant correlations between participants´ reports on how truthful others consider they are and their real honesty; however, future research should replicate this finding by correlating participants´ behavioral measures of honesty with ratings provided by the close acquaintances themselves.

An additional objective of the present study was to examine whether facial babyishness and attractiveness were independent from each other. Although Berry (1991b), in an excellent study where a large sample of facial photographs was used, found evidence that attractiveness and babyfacedness were independent dimensions, our data revealed a marginally significant association between facial babyishness and attractiveness. Almost all our stimulus people were women. In general, women are more babyfaced than men (Enlow, 1982), and research shows that gender-prototypical facial features are perceived as more attractive than the gender-discrepant (e.g., Langlois & Roggman, 1990; Zebrowitz, 1997; see however Perrett, May, & Yoshikawa, 1994; Wheeler, cited in Berry, 1991b). This suggests that positive correlations between babyfacedness and attractiveness should be found for women, and this is what our data reveal, although the tendency does not reach the .05 significance level. Also, it would be reasonable to expect negative correlations between males´ facial babyishness and their attractiveness. Since we had only five males in the sample of stimulus persons we could not test this prediction.

Finally, we looked at students´ tendency to deceive or their reservations about lying. Most students were willing to participate in most experimental procedures, either truthful or deceptive. In fact, there were no significant differences between the proportion of participants who agreed to collaborate in either of the two nondeceptive experimental procedures and those who agreed to participate in a number of deceptive studies, namely the false feedback, bogus reward, rigged contest, and public speech experiments. Actually, the students were willing to participate in 4.40 out of 7 deceptive studies although, fortunately, the strength of their willingness to do so was moderate (2.51 in a 0 – 6 scale). Although our requirement to accept to collaborate in at least two studies could have somewhat increased these figures (although participants could have chosen the two nondeceptive studies), it is surprising that as many as a 63.5 % of the students wrote the deceptive letter, though it was emphasized that they could freely choose not to do so and, indeed, the emotional consequences for the recipient of the letter could be devastating if he or she discovered the deception. In addition, the reason most frequently mentioned by those who did not write the letter was that they considered they were not skilled enough to lie convincingly –only a few argued they considered deception unethical, or possibly damaging for the recipient. This indicates that our students do not have very stringent moral standards. It seems the main reason why some of them did not write the deceptive note was their perception of low self-efficacy for lying in that situation. Otherwise, they would have lied. Our results are in line with other studies examining the role of self-efficacy (e.g., Bandura, 1977, 1997) in the performance of socially rep-

rehensible acts. For example, recent research conducted by Garrido (2000) indicates that perceptions of self-efficacy for engaging in delinquent behavior are a strong predictor of delinquency.

As said earlier, research on social perception of faces is relevant for other areas of inquiry, such as deception detection research. It has been suggested facial stereotypes could account for the demeanor bias (e.g., Bond *et al.*, 1994; Masip *et al.*, 1999; Zebrowitz *et al.*, 1996). However, observers´ agreement as to which individuals look honest and which look dishonest (regardless of the accuracy of those stereotypes) may not be necessarily reflected in real decisions made in specific situations. That is to say, the fact that face A is considered by everybody to be dishonest, and face B is considered by everybody to be honest, does not imply that the same statement (a given statement) will be judged as deceptive when attributed to face A and as truthful when attributed to face B. Scholars ask research participants to judge photographs in terms of a series of personality dimensions (e.g., honesty), and clear tendencies (e.g., that there is consensus in considering face A as dishonest and face B as honest) are found. But actually, when other information is available, when people are not asked to judge the *person* (i.e., "is he/she honest or deceptive?") but his or her *behavior* (i.e., "is this statement [delivered by that person] truthful or deceptive?"), is person A judged to be deceptive more often than person B, and person B judged to be truthful more often than person A? If so, and if social stereotypes are *not* accurate, the issue would be worrisome, because then wrong impressions based on someone´s facial appearance would make us behave in a given way towards that person (e.g., considering he or she is lying regardless of what he or she says – the demeanor bias–), and our behavior could have serious consequences for him or her (e.g., he or she being acquitted or condemned in court if we are jurors). Ongoing research is examining these questions.

Acknowledgements

The authors are very grateful to Gema Martín and Ana María Calvo, who participated as research assistants in this study. The research reported here was supported by the *Junta de Castilla y León*, *Programa de Apoyo a Proyectos de Investigación*, Ref. SA52/00B.

References

Alley, T. R. (1988). Physiognomy and social perception. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces* (pp. 167-186). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Alley, T. R., & Hildebrandt, K. A. (1988). Determinants and consequences of facial aesthetics. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces* (pp. 101-140). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences. A meta-analysis. *Psychological Bulletin, 111*, 256-274.

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In M. H. Guetzkow (Ed.), *Groups, leadership and men.* (pp. 117-190). Pittsburgh: Carnegie.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1997). *Self-efficacy, the exercise of control.* New York: W. H. Freeman & Company.

Bem, D. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 6*, 1-62.

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition, 8*, 343-361.

Berry, D. S. (1991a). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*, 298-307.

Berry, D. S. (1991b). Attractive faces are not all created equal: Joint effects of facial babyishness and attractiveness on social perception. *Personality and Social Psychology Bulletin, 17*(5), 523-531.

Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Bulletin, 15*(2), 266-279.

Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312-323.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100*(1), 3-18.

Berry, D. S., & McArthur, L. Z. (1988). The impact of age-related craniofacial changes on social perception. In T.

R. Alley (Ed.), *Social and applied aspects of perceiving faces*. (pp. 63-87). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 7*, 157-215.

Bond, C. F. Jr., & Robinson, M. (1988). The evolution of deception. *Journal of Nonverbal Behavior, 12*(4), 295-307.

Bond, C. F. Jr., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15*(4), 523-534.

Bond, C. F. Jr., Kahler, K. N., & Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology, 21*, 331-345.

Bond, C. F. Jr., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14*(3), 189-204.

Bond, C. F. Jr., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology, 63* (6), 969-977.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546-553.

Brownlow, S. (1993). Seeing is believing: Facial appearance, credibility, and attitude change. *Journal of Nonverbal Behavior, 16*(2), 101-115.

Brownlow, S., & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior, 14*(1), 51-60.

Bruce, V., & Young, A. (1998). *In the eye of the beholder. The science of face perception*. Oxford: Oxford University Press.

Bull, R. (1982). Physical appearance and criminality. *Current Psychological Reviews, 2*, 269-282.

Bull, R., & Green, J. (1980). The relationship between physical appearance and criminality. *Medical Sciences and the Law, 20*(2), 79-83.

Bull, R., & McAlpine, S. (1998). Facial appearance and criminality. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and law. Truthfulness, accuracy and credibility* (pp. 59-76). London: McGraw-Hill.

Caro-Baroja, J. (1987). *La cara, espejo del alma. Historia de la fisiognómica*. Madrid: Galaxia Gutenberg-Círculo de Lectores.

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). North-Holland: Elsevier Science Publishers.

Cooper, H., & Good, T. (1983). *Pygmalion grows up: Studies in the expectation communication process*. New York: Longman.

Darley, J. M., & Fazio, R. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35*, 867-881.

della Porta, G. B. (1644). *Della fisionomia dell´huomo*. Venice.

della Rocca, B. (1536). *Naturalis philosophiæ ac medicinæ doctoris, physiognomiæ & chiromantiæ compedium*. Strasburg: Joannem Albertum.

Deutsch, M., & Gerard, H. (1955). A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology, 51*, 629-636.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285-290.

Ekman, P. (1992). *Telling lies. Clues to deceit in the marketplace, politics, and marriage*. (2nd. ed.). New York: W. W. Norton & Company.

Ekman, P. (1997). Deception, lying, and demeanor. In D. F. Halpern & A. E. Voiskounsky (Eds.), *States of mind. American and post-Soviet perspectives on contemporary issues in psychology*. (pp. 93-105) New York: Oxford University Press.

Enlow, D. H. (1982). *Handbook of facial growth*. (2nd. edition). Philadelphia: Saunders.

Fancher, R. E. (1988). Gall, Flourens and the phrenological movement. In B. Ludy (Ed.), *A history of psychology: Original sources and contemporary research*. (pp. 101-108). New York: McGraw-Hill.

Galton, F. (1883). *Inquires into human faculty and its development*. London: Macmillan.

Garrido, E. (2000). *Explicación del comportamiento moral e inmoral en la Teoría Social Cognitiva*. Unpublished manuscript.

Geis, F. L., & Moon, T. H. (1981). Machiavellianism and deception. *Journal of Personality and Social Psychology, 41*(4), 766-775.

Keating, C., Mazur, A., & Segall, M. (1981). A cross-cultural exploration of physiognomic traits of dominance and happiness. *Ethology and Sociobiology, 2*, 41-48.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin, 116*, 245-258.

Kenny, D. A.,, Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88-97.

Kretschmer, E. (1921). *Köperbau und charakter*. Berlín: Springer-Verlag.

Langlois, J. H. (1986). From the eye of the beholder to behavioral reality: The development of social behaviors and social relations as a function of physical attractiveness. In C. P. Herman, M. P. Zanna, & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario symposium, Vol. 3*. (pp. 23-51). Hillsdale, NJ: Lawrence Erlbaum.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*, 115-121.

Lavater, J. C. (1793). *Essays on physiognomy: For the proportion of the knowledge and the love of mankind*. London: Robinson.

Levesque, M. J., & Kenny, D. S. (1993). Accuracy of behavioral predictions at zero acquaintance: A social re-

lations analysis. *Journal of Personality and Social Psychology, 65*, 1178-1187.

Lombroso, C. (1895). *L´homme criminel*. Paris.

Masip, J., Garrido, E., & Herrero, C. (1999). *Factors contributing to police officers´ and lay people´s credibility judgments: The moment observers make their decision, and the senders´ visible and auditory behavior and facial appearance*. Manuscript submitted for publication.

McArthur, L. Z., & Apatow, K. (1983-1984). Impressions of baby-faced adults. *Social Cognition, 2*(4), 315-342.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215-238.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology, 37*, 233-256.

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

Miller, G. R., deTurck, M. A., & Kalbfleisch, P. J. (1983). Self monitoring, rehearsal, and deceptive communication. *Human Communication Research, 10*, 97-117.

Montepare, J. M. , & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. In M. Zanna (Ed.), *Advances in experimental social psychology, 30*, 93-161.

Moreau, M. (Ed.), (1820). *L´art de connaître les hommes par la Physionomie par Gaspard Lavater. Nouvelle édition, corrigée et disposée dans un ordre plus méthodique; précédée d´une Notice historique sur l´Auteur; augmentée d´une Exposition des recherches ou des opinions de La Chambre, de Porta, de Camper, de Gall, sur la physionomie; d´une Histoire anatomique et physiologique de la face, etc.* (10 vols.) Paris: Depélafoi, Libraire.

Moskowitz, D. S. (1990). Convergence of self-reports and independent observers: Dominance and friendliness. *Journal of Personality and Social Psychology, 58*, 1096-1106.

Moulton, J., Blake, R. & Olmstead, J. (1956). The relationship between frequency of yielding and the disclosure of personal identity. *Journal of Personality, 24*, 339-347.

Mueller, J. H., Thompson, W. B., & Vogel, J. M. (1988). Perceived honesty and face memory. *Personality and Social Psychology Bulletin, 14*(1), 114-124.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Attractive characteristics of female faces: Preference for non-average shape. *Nature, 368*, 239-242.

Riggio, R. E., & Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology, 45*(4), 899-915.

Riggio, R. E., Tucker, J., & Throckmorton, D. (1987). Social skills and deception ability. *Personality and Social Psychology Bulletin, 13*, 568-577.

Riggio, R. E., Tucker, J., & Widaman, K. F. (1987). Verbal and nonverbal cues as mediators of deception ability. *Journal of Nonverbal Behavior, 11*(3), 126-145.

Rohrer, J. H., Baron, S. H., Hoffman, E. L., & Swander, D. V. (1954). The stability of autokinetic judgments. *Journal of Abnormal and Social Psychology, 49*, 595-597.

Sheldon, W. H. & Stevens, S. S. (1942). *The varieties of temperament*. New York: Harper.

Sheldon, W. H., Stevens, S. S., & Tucker, W. B. (1940). *The varieties of human physique*. New York: Harper.

Sheriff, M. (1936). *The psychology of social norms*. New York: Harper.

Snyder, M. (1984). When beliefs create reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 18*, 248-306.

Viguera, M. J. (1977). *Dos cartillas de fisiognómica*. Madrid: Madrid.

Vrij, A. (1992). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of Social Psychology, 133*(5), 601-610.

Vrij, A. (1998). Nonverbal communication and credibility. In A. Memon, A. Vrij, & R. Bull. (Eds.), *Psychology and law. Truthfulness, accuracy and credibility.* (pp. 32-58). New York: McGraw-Hill.

Vrij, A., & Winkel, F. W. (1993). Objective and subjective indicators of deception. *Issues in Criminological and Legal Psychology, 20*, 51-57.

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior, 21*(2), 87-102.

Yates, D. H. (1967). La seudopsicología. In E. Brunswick, R. F. Creegan, & D. H. Yates, *Los marcos de referencia en psicología*. Buenos Aires: Paidós.

Zebrowitz, L. A. (1997). *Reading faces. Window to the soul?* Boulder, Colorado: Westview Press.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology, 28*(6), 1143-1152.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don´t all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology, 65*, 85-101.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin, 22*(12), 1258-1269.

Znakov, V. (1997). Comprehension of lies. A Russian view. In D. F. Halpern & A. E. Voiskounsky (Eds.), *States of mind. American and post-Soviet perspectives on contemporary issues in psychology.* (pp. 79-91) New York: Oxford University Press.

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.

(Artículo recibido: 29-12-2000, aceptado: 25-5-2001)