Journal of Personality and Social Psychology
2004, Vol. 86, No. 5, 668–679

0022-3514/04/$12.00 DOI: 10.1037/0022-3514.86.5.668

Encoding Under Trust and Distrust: The Spontaneous Activation of Incongruent Cognitions

Yaacov Schul and Ruth Mayo
Hebrew University

Eugene Burnstein
University of Michigan

Past studies of strategic thinking have shown that the encoding of the message information becomes more complex under distrust. Receivers process the information as if they are trying to protect themselves from being misled by testing alternative potential interpretations. The present study investigates the possibility that when people are mistrustful they spontaneously activate associations that are incongruent with the given message. Findings from 3 experiments suggest that, even when the distrust is unrelated in any meaningful way to the message and even when receivers are unable to prepare a strategic response, the cognitive system reacts to distrust by automatically inducing the consideration of incongruent associations—it seems designed to ask, "and what if the information were false?" The theoretical implications of the results for theories of social perception and persuasion are discussed.

Deception and its complement, distrust, are among the most perplexing and pervasive challenges of social life. Evolutionary theory assumes they are inherent to living in groups and a key problem of adaptation (Griffin, 1992). The social psychological literature on the "epidemiology" of deception and distrust offers much empirical support at least in respect to prevalence. For example, DePaulo, Kashy, Kirkendol, Wyer, and Epstein (1996) found that participants in a community survey said they lied in about one of five social interactions, and college students reported doing so in one of three (see also, DePaulo & Kashy, 1998; Feldman, Forrest, & Happ, 2002). Distrust and suspicion seem no less common within and among organizations (Kramer, 1999; Lewicki, McAllister, & Bies, 1998); it is well known that consumers often distrust information put out by these organizations, especially when it is about their products (e.g., Dyer & Kuehl, 1978; Tellis, 1997). The view that "politicians pander" is not uncommon (McGraw, Lodge, & Jones, 2002), and even journalists, admitting that business pressures unduly influence what they write, have become concerned that readers are growing to distrust them (Pew Research Center for the People and the Press, 1999). Hence, it seems reasonable to conclude that in many, if not most, of their dealings with others, people are aware of the possibility of being misled (DePaulo et al., 1996; Schul & Burnstein, 1998). Our research examines how the human cognitive system has adapted to this sort of awareness. Specifically, it explores whether a feeling of distrust leads to a different form of information processing than a feeling of trust.

Coping with distrust would be easy if people could recognize liars. After many thousands of generations of living in groups, one might think that human beings would have evolved into highly accurate social perceivers. Yet, as dozens of studies suggest, the accuracy of interpersonal perception is modest at best, attesting to the complexity of the task people face when trying to decipher others' inner states (Funder, 1995). The reason, of course, is that as our social coding skills have evolved, so too have our skills in masking thoughts and feelings. The result is a Darwinian "arms race" between senders attempting to deceive and receivers striving to detect deception. In this contest, accuracy seems to succumb more often than not. There is ample evidence that competency in unmasking deception is rare (Anderson, DePaulo, Ansfield, Tickle, & Green, 1999; DePaulo & Friedman, 1998; Ekman & O'Sullivan, 1991) and, worse, that people are unaware of their incompetence. For example, DePaulo, Charlton, Cooper, Lindsay, and Muhlenbruck (1997) showed that the confidence receivers have in their accuracy is virtually unrelated to their actual accuracy.

Although receivers are not adept at separating cheaters from truthful communicators, they can still protect themselves from falsehoods to some extent by encoding incoming messages in a suspicious context differently from those in a trustworthy context. Discriminative encoding of this sort is suggested by Schul, Burnstein, and Bardi (1996), who found that when respondents had to read a set of messages about a person, those who were made suspicious about the validity of one message needed more time to read the set and integrate the messages than was needed by those who were not made suspicious. This was interpreted to mean that suspicion increased the complexity of encoding by delaying the "freezing" or integration of the messages within a single interpretative frame while promoting the construction and testing of alternative interpretations, called *counter-scenarios* (cf. Kruglanski,

Yaacov Schul and Ruth Mayo, Department of Psychology, Hebrew University, Jerusalem, Israel; Eugene Burnstein, Department of Psychology, University of Michigan.

The research reported here was supported by grants from the United States–Israel Binational Science Foundation. We thank Tal-Chen Strauss-man and Hila Goldberg for their help in running the experiments.

Correspondence concerning this article should be addressed to Yaacov Schul, Department of Psychology, Hebrew University, Mount Scopus, Jerusalem 91905 Israel. E-mail: yschul@mscc.huji.ac.il

1989; Kruglanski & Freund, 1983). This analysis suggests that receivers who suspect the validity of messages they are about to receive encode the messages as if they were true and, at the same time, as if their opposite was true. A similar conclusion has been suggested by Fein and colleagues (Fein, Hilton, & Miller, 1990; Fein, McCloskey, & Tomlinson, 1997; Hilton, Fein, & Miller, 1993). They found that when individuals suspect a possible hidden motive that may account for the actions of a protagonist, they process information as if they are examining the protagonist's actions in two opposite scenarios: one consistent with the explicit motive given in the story and the other consistent with the hidden motive.

If one cannot trust another's word, it makes good sense not only to probe a little more but also to consider what might happen if the opposite of what the other says is true. Thus, on suspecting a communicator, receivers can deliberately attempt to protect themselves by thinking about potential counter-scenarios. Although such attempts might be cognitively taxing (Schul et al., 1996), it has been shown that an encoding strategy of this sort facilitates successful discounting of invalid messages and reduces various biases such as message-order effects or the correspondence bias (Fein, 1996; Fein et al., 1997; Schul, 1993; Schul et al., 1996).

The present study investigates the possibility of a variant of counter-scenario processing that is not strategic or deliberate. We explore the possibility that under distrust people spontaneously activate associations that are incongruent with (or opposite to) the given message. This would make processing under conditions of distrust quite different than processing under trust. Specifically, there is a good deal of evidence showing that perceivers tend initially to accept messages, that is, treat them for the moment as if they were true. In case the message is not valid, perceivers can and do modify their initial coding at some later point in processing, given that they have sufficient motivation, ability, and cognitive resources (Gilbert, Pelham, & Krull, 1988; Newman & Uleman, 1989; Trope, 1986). The tendency to initially accept the given information has been shown in persuasion research (Festinger & Maccoby, 1964; Petty, Wells, & Brock, 1976), hypothesis testing (Snyder & Swann, 1978; Wason & Johnson-Laird, 1972), and the processing of denials (Wegner, Coulton, & Wenzlaff, 1985; Wegner, Wenzlaff, Kerker, & Beattie, 1981). As has been suggested by Gilbert, Tafarodi, and Malone (1993), "belief is first, easy, and inexorable [whereas] . . . doubt is retroactive, difficult, and only occasionally successful" (p. 231).

Although initial acceptance might be a desirable processing strategy in the context of trust, it is not functional in the context of distrust. Indeed, our study investigates whether information processing follows a more adaptive strategy under conditions of distrust—namely, whether in this case receivers activate ideas opposite to or incongruent with those in the message. Hence, when politicians promise to keep their word in an election campaign, skeptical receivers might immediately encode the promise in a state of disbelief and, thus, spontaneously think about instances of broken promises. Similarly, when someone who is trying to sell something urges us to make up our mind because it is his or her last one or because the price will double tomorrow, we think that the person may very well have a full stock of it and that the price may stay the same or even drop tomorrow. Thus, analogous to counter-scenario generation, we propose that under conditions of distrust a message spontaneously activates alternative meanings incongruent with it.

Our research compares the extent to which message-incongruent associations are activated under conditions of trust with what happens under conditions of distrust. It is important to note that in the experiments reported later the trust/distrust manipulation is completely unrelated to the message. This allows us to explore whether the mere state of distrust is sufficient to induce thinking about incongruent alternatives. Experiments 1 and 3 use an identical trust/distrust manipulation and different dependent measures. In these experiments, trust or distrust is signaled by faces. In Experiment 1 each face was followed by a prime word and a target word. We investigate whether trustworthy faces facilitate processing of congruent prime–target pairs whereas untrustworthy faces facilitate processing of incongruent prime–target pairs. In Experiment 3 respondents generate an association for a word that appears together with a trustworthy or an untrustworthy face. An assessment is made of the extent to which associations elicited in the context of trust are different from those elicited in the context of distrust. Experiment 2 makes use of a different manipulation of trust/distrust, one that does not involve faces. However, it uses the same dependent measure as Experiment 1, giving our findings a greater degree of generalizability.

Experiment 1

Faces, it is said, are a window to the soul. People take this seriously; not only do they believe they can infer personality traits from other people's faces, but they also evidence an impressive degree of consensus in their inferences (Berry & Waro-Finch, 1993; Berry & McArthur-Zebrowitz, 1986; Hassin & Trope, 2000). Not surprisingly, they believe that being keen analysts of faces allows them to detect deception. Akehurst, Koehnken, Vrij, and Bull (1996), for example, asked police officers (presumably more knowledgeable in detecting deception) and lay people (presumably less knowledgeable) about their beliefs regarding 64 cues for detecting deception. Overall, the beliefs of police officers and laypersons were similar. As expected, different aspects of facial posture and facial expression (e.g., change in the line of sight, eye contact, eye blinks, lip biting) were believed to leak information about untrustworthiness (see Anderson et al., 1999, for the null relationship between the beliefs and the actual diagnosticity of different cues). Our study capitalizes on such beliefs about the role of the face in revealing deception by associating trust and distrust with different facial cues.

The experimental session started with an induction phase designed to link the perception of trustworthiness to eye shape. More specifically, during induction respondents were presented with 20 different faces with round eyes and 20 different faces with narrow eyes. The round-eyed faces were paired with correct messages and the narrow-eyed faces with false messages. It has been shown that whereas round eyes are perceived as a sign for trust, narrow eyes raise suspicion (for a review, see Zebrowitz, 1997, pp. 84–85). Therefore, we assume that such a pairing reinforces through implicit learning (e.g., Hill, Lewicki, Czyzewska, & Schuller, 1990) a preexisting association between eye shape (round or narrow) and the trustworthiness of the face. We tested this assumption in a pretest and found that, following the induction phase, respondents did judge round-eyed faces as more trustworthy than narrow-eyed ones.

Phase 2 of the experiment tested whether distrust elicits different associative processes than trust. Specifically, in each trial of the second phase, respondents saw a new face (with either narrow or round eyes) and a pair of words: a prime and a target. Their task was to indicate whether the target word was a noun or an adjective. We measured the time respondents took to make this judgment. According to our theoretical analysis, the nature of the face determines the influence that the prime word has on the target word. When the face signals trust (i.e., round eyes), respondents are assumed to spontaneously activate congruent associations, at least more so than when the face signals distrust (i.e., narrow eyes). Conversely, a context of distrust should facilitate incongruent associations more than a context of trust. As these are opposite patterns of facilitation, we should observe an interaction between the kind of face (trustworthy vs. distrusted) and the degree of facilitation (incongruent vs. congruent associations).

There are several noteworthy features of the task. First, the faces are irrelevant for determining whether the target word is an adjective or a noun. This allows us to test whether the presence of trust or distrust by itself influences processing. Second, the prime word appeared for a very short period (less than 100 ms) and was immediately replaced by the target word. This allowed respondents to perceive the former above awareness, but not to use controlled or strategic processes to benefit from the prime–target relationship (e.g., Neely, 1977; see also, Wegner & Bargh, 1998). Finally, because we measure the time respondents take to classify the target word as an adjective or a noun, it is unlikely that they could have formulated hypotheses that might link the kind of face (round-eyed vs. narrow-eyed) to the facilitation of congruent and incongruent targets. Thus, if the faces do influence the nature of activation in this situation, the influence is not likely to be a reflection of respondents' expectations or plans.

Finally, Experiment 1 included two control conditions. A no-face condition, in which faces were replaced by geometrical figures, allowed us to test whether the pattern of priming is influenced by the presence of faces per se. A no-prime condition, whereby the target word followed a nonword string, allowed us to examine whether the faces by themselves influence how the target words are perceived.

Method

Participants

Fifty-four students at the Hebrew University (Jerusalem, Israel) participated in the experiment, 28 in the experimental condition, 14 in the no-prime control condition, and 12 in the no-face control condition. Seven of these students (4 from the experimental condition, 2 from the no-prime condition, and 1 from the no-face condition) had more than 20% errors in the priming task and were eliminated from the analyses. Nine additional respondents participated in the pretests for testing the trustworthiness of the round- versus narrow-eyed faces. Respondents received either payment or extra credit for their participation.

Procedure

Phase 1 (induction). This phase of the experiment created a link between the round eyes and trust and between the narrow eyes and distrust. To this end, respondents alternated between exposure and memory trials. During the exposure trials respondents saw on the computer screen a face and a trivia sentence, which was attributed to the owner of the face (see top panel of Figure 1). They had two tasks: (a) to determine whether the sentence was true or false and (b) to remember "who said what." After indicating their answer, respondents were told whether the sentence was actually true or false. Unbeknownst to the respondents, narrow-eyed faces always appeared with false trivia sentences and round-eyed faces with true trivia sentences. After exposure to 20 faces and trivia sentences, respondents were tested on their memory for "who said what." In each of eight memory trials, they were presented with a face and a trivia sentence and asked whether that person had said the sentence. This procedure was used to direct the respondents' attention to the facial cues, thus allowing them to learn the correspondence between the type of eyes and the truth of the trivia sentences.

Overall, following two exposure–memory blocks, respondents were presented with 20 narrow-eyed faces that were paired with false trivia sentences and 20 round-eyed faces that were paired with true trivia sentences. After going through these blocks once, respondents were informed that their performance was suboptimal; hence, the blocks were repeated using a different order. We assume (see following text) that this procedure of induction reinforces the preexisting link between the nature of the eyes and the degree of trust.

Phase 2 (priming). Each trial in this phase started with the presentation of a face for 800 ms (see bottom panel of Figure 1). Then a prime word was superimposed on the face a little below the eyes. After being shown for 82 ms, the prime word was replaced by a target word. The target word remained on the screen until a response was made or until a 2-s response window had ended. Respondents were to indicate whether the target word was an adjective (pressing the key *l*) or a noun (pressing the key *a*). They were informed that we were measuring their speed of response and their accuracy. Respondents were encouraged to be as fast and as accurate as possible. After each response they were given feedback as to whether they were correct. Stimulus presentation and response collection were controlled by Inquisit (Inquisit 1.32, 2002).

Stimuli Construction

Faces were generated by *Faces* (Faces 3.0, 1998), a computer program that allows the generation of composite images from individual features. We generated 80 different faces, 40 with round eyes and 40 with narrow eyes. The 80 faces were divided into two sets, with 40 faces (20 round- and 20 narrow-eyed) used in the induction phase and the other 40 in the priming phase (half of the pictures in each set were female faces).

The trivia sentences were generated from a variety of sources. We ended up with four sets: 10 true–easy (e.g., "The capital of Greece is Athens"), 10 true–hard (e.g., "Nemesis is the goddess of revenge"), 10 false–easy (e.g., "Steven Hawking is a painter"), and 10 false–hard (e.g., "The first traffic light was installed in England prior to the first world war"). The classification of trivia sentences into easy and hard was based on a pretest with 60 respondents.

Prime–target pairs were constructed in the following way. We generated a set of 40 adjective triads, each consisting of a target adjective word and two adjective primes, one semantically congruent with the target and the other semantically incongruent with it (e.g., *transient–temporary* or *transient–permanent, hollow–empty* or *hollow–full*)[1]. In addition, we generated 40 adjective–noun pairs. In these pairs the prime was an adjective and the target was a noun. As the prime word was always an adjective, it could not help the participants in determining whether the target was an adjective or noun.

[1] We used Hebrew words. The examples are free translations.

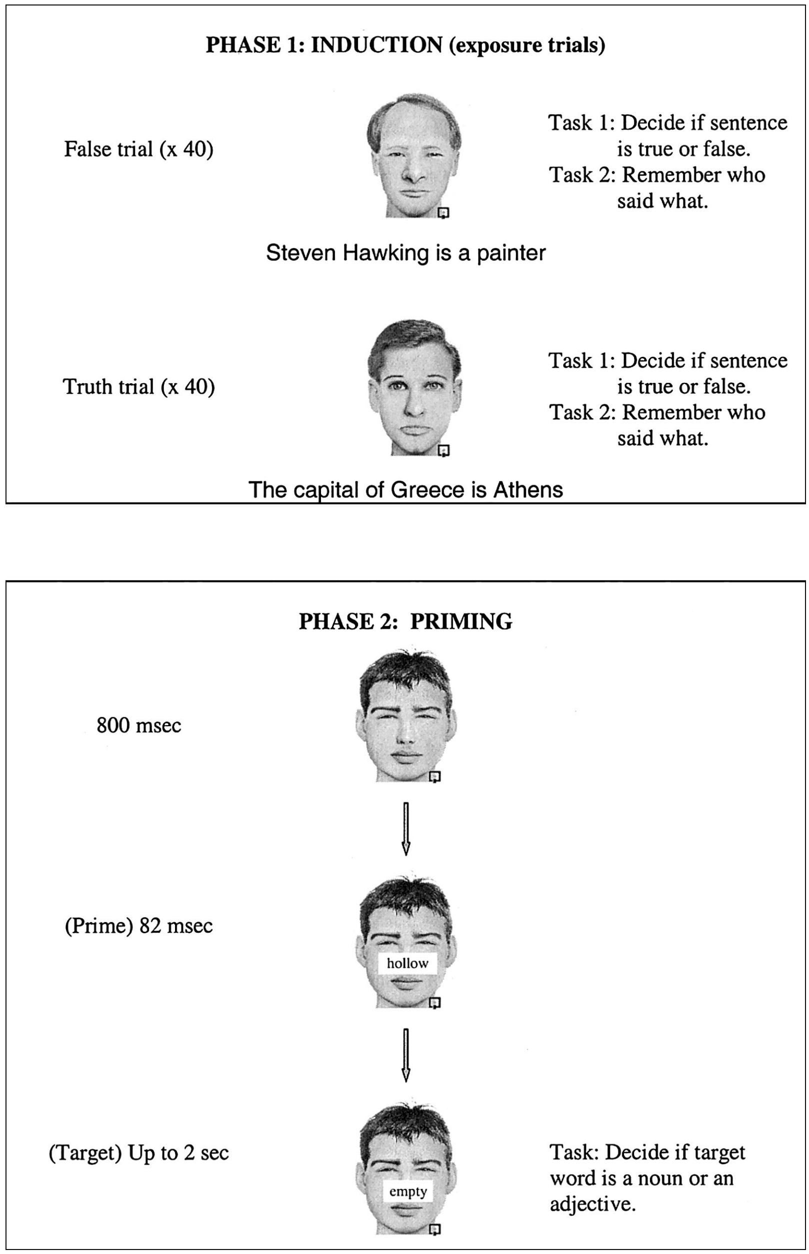

Figure 1. Procedures of Experiment 1. msec = milliseconds; sec = seconds.

In the no-face control condition the faces were replaced by geometrical figures. We used four different polygons, each of which had five different grids. In the no-prime control condition we replaced the prime word by a single meaningless string of letters.

Counterbalancing

Each respondent saw 40 adjective–noun pairs and 40 adjective–adjective pairs. The first word in each pair served as a prime, the second as a target. Pairs were selected with the restriction that the 80 pairs included 80 different target words. The 40 adjective–adjective pairs consisted of 20 congruent pairs (e.g., *transient–temporary*) and 20 incongruent pairs (e.g., *hollow–full*). Ten of the congruent pairs appeared with trustworthy (round-eyed) faces and 10 with untrustworthy (narrow-eyed) faces. Similarly, 10 incongruent pairs appeared with trustworthy faces and 10 with untrustworthy faces. Half of the adjective–noun pairs appeared with trustworthy faces and half with untrustworthy faces.

On a more technical level, each respondent saw 80 prime–target pairs (40 adjective–noun and 40 adjective–adjective), but only 40 faces, meaning that each face appeared twice during Phase 2. The 80 pairs were split into two blocks of 40 trials, and faces that appeared with an adjective–noun pair in Block 1 appeared with an adjective–adjective pair in Block 2. The 40 adjective–adjective pairs were divided into eight sets, each containing five pairs. The 40 adjective–noun pairs were divided similarly. The first four adjective–adjective and adjective–noun sets comprised Block 1. The rest were considered Block 2. These were ordered differently for each participant, with the constraint that for every participant who saw a particular pattern of congruent and incongruent pairs, there was another participant who saw the exact mirror image (i.e., incongruent and congruent pairs). The exact combinations of faces and pairs, as well as the order of presentation within each set, varied randomly between respondents.

Finally, each block started with two buffer trials with different faces and different prime–target pairs. The results from the buffer trials were not included in the statistical analyses.

Design

Phase 1 of the experiment was identical for all respondents. The different conditions varied in the nature of the stimuli presented in Phase 2. The experimental condition consisted of a 2 (round-eyed vs. narrow-eyed faces) × 2 (congruent vs. incongruent prime–target) within-subject design. In addition, two control conditions were run, both of which were variations on the experimental condition. In the no-face condition geometrical figures replaced the faces. In the no-prime condition a single nonword string replaced the primes.

Pretesting of the Induction Procedure

To test whether we were successful in linking round-eyed faces to trust and narrow-eyed faces to distrust, we ran the following pretest: Nine respondents went through the induction phase. They were then presented with the 40 faces that were to be used in the priming phase of the experiment proper. They were instructed to rate each face for its trustworthiness on an 8-point scale. Each of the 9 respondents rated the round-eyed faces as more trustworthy than the narrow-eyed ones. An analysis of variance (ANOVA) indicated that narrow-eyed faces were rated as significantly less trustworthy than round-eyed faces (3.34 vs. 5.29), $F(1, 8) = 38.35$, $p < .01$.

Results

Data Preparation

Prior to the statistical analyses, we performed several preparatory steps. First, each respondent saw 80 target words. He or she was to indicate whether the target was an adjective or a noun within a 2-s response window. Participants failed to respond on time in 22 trials (less than 1% of the trials). These trials were eliminated from all analyses. Also, we eliminated response latencies from 84 trials that were more than 3 standard deviations from the mean (about 2%). Finally, response accuracy was examined. Overall accuracy was high (93% in the experimental condition, 94% in each of the two control conditions). In the analyses of response latency, trials with incorrect responses were excluded. Thus, for each respondent, we computed the mean latency of responding correctly to target words in the different face–prime–target combinations. The means appear in Table 1.

Analyses of Priming

Respondents in the experiment were presented with face–prime–target combinations that included either noun targets or adjective targets. As we were testing the hypothesis that the effect of priming would differ as a function of the trustworthiness of the face, and as there was no relationship between the prime and the noun targets, our main analyses were focused on the trials in which the target was an adjective. Thus, the hypothesis that type of face and the congruency of the prime–target relationship interact in determining the speed of response in Phase 2 is tested by the interaction contrast that compares the differences between congruent and incongruent primes following trustworthy faces versus untrustworthy faces.

Consider first the responses in the no-prime control condition. This control condition is important because it allows us to assess whether the nature of the face influenced the response to the target word when there was no prime. The speed of the participants' reactions did not vary as a function of the face, neither for the adjectives nor for the nouns. A two-way repeated-measures ANOVA (Face × Target) revealed that neither the main effect of face, $F(1, 11) < 1$, nor the Face × Target interaction, $F(1, 11) < 1$, was significant. Thus, it appears that, by themselves, the two types of faces did not induce different kinds of processing.

Next, we consider responses in the no-face control condition. Analysis of these responses is informative about the differences in the associative strength of the congruent and incongruent pairs when they are not influenced by the presence of trustworthy or untrustworthy faces. A one-way repeated-measure ANOVA revealed that although participants reacted to the incongruent targets a little faster than to the congruent target (678 vs. 697), this difference failed to reach significance, $t(11) = 1.27$, $p > .20$. This suggests that our construction of triads was successful in the sense that the congruent and the incongruent primes were not very

Table 1
Means (and Standard Deviations) of the Latency of Accurate Responses (in Milliseconds) and Proportion of Correct Responses From the Adjective–Noun Classification Task (Experiment 1)

Condition	Congruent prime adjective target	Adjective target	Noun target
No prime		No prime	No prime
Untrustworthy faces		790 (85)/.93	753 (77)/.95
Trustworthy faces		785 (79)/.93	750 (71)/.95
No face		Incongruent prime	Irrelevant prime
Polygons	697 (67)/.95	678 (74)/.91	678 (41)/.95
Experimental			
Untrustworthy faces	775 (138)/.93	760 (128)/.96	749 (107)/.94
Trustworthy faces	754 (137)/.94	783 (125)/.93	750 (99)/.92

different in their capacity to prime the target words. Comparison of the two control conditions indicates, not surprisingly perhaps, that the presence of faces slowed down responses to the target words relative to the presence of prime words.

Although we found no evidence that the faces influenced the speed of processing of the target words per se, the results from the experimental condition show that they did influence the pattern of priming, as hypothesized. To test this hypothesis, we performed a two-way repeated-measures ANOVA on the speed with which the targets were identified. As hypothesized, the theoretically relevant interaction contrast was significant, $F(1, 23) = 5.15$, $p < .05$. It suggests that the facilitation of congruent and incongruent primes differed under trust and distrust. A simple effect analysis revealed that, in line with our theoretical predictions, congruent primes facilitated the processing of the target adjectives significantly more following the round-eyed faces (trust) than following the narrow-eyed faces (distrust), $t(24) = 1.81$, $p < .05$, one-tailed. Incongruent primes, in contrast, facilitated the processing of the target adjectives significantly more following distrusted faces, $t(24) = 2.03$, $p < .05$, one-tailed. Finally, the trustworthiness of the face did not influence the processing of the nouns, which is not surprising, as the prime words were irrelevant to the nouns (cf. the no-prime control condition).

Performance Accuracy

Table 1 also depicts the mean performance accuracy. Examination of the entries in the experimental condition reveals small differences among the six cells. In particular, the 2 (trustworthy vs. untrustworthy face) $\times$ 2 (congruent vs. incongruent prime) interaction contrast failed to reach significance, $F(1, 23) = 1.61$, $p > .2$. Note that the conditions that produced the fast responses (trustworthy face + congruent prime and untrustworthy face + incongruent prime) led to greater levels of accuracy. Thus, the greater performance speed in these cells does not reflect a speed–accuracy trade-off. The no-prime condition showed also considerable similarity in the accuracy level between the different cells. Interestingly, the analysis of the no-face control condition revealed that responses in case of incongruent primes were less accurate than those for the congruent primes (.91 vs. .95), $F(1, 10) = 13.43$, $p < .01$. This might be related to the (nonsignificant) faster processing speed of the former.

Discussion

Participants in Experiment 1 had to determine whether a target word was an adjective or a noun. The target word followed a prime word, which was presented for a brief period that allowed for comprehension but not for initiation of any willful or deliberate processing strategies. When the prime–target pair was presented in the context of a trustworthy face, the targets benefited from congruent primes more than from incongruent primes. However, in the context of an untrustworthy face, the targets benefited more from incongruent than from congruent primes. This sort of pattern is consistent with past research demonstrating that suspicion about the validity of a message leads individuals to access alternative scenarios. Experiment 1 shows that the extent of priming facilitation of congruent and incongruent words was determined by the trust/distrust of the face that preceded the prime, so that in the presence of an untrustworthy face, we observed faster activation of incongruent than congruent associations, even when the face was irrelevant to the task.

We believe that the activation of incongruent associations was not the result of a conscious processing strategy, because the prime–target interval was very short (less than 100 ms; see Neely, 1977) and because the respondents were unlikely to come up with a theory linking the shape of the eyes to the facilitation of congruent and incongruent targets. Rather, we think that the observed effect has to do with a generalized pattern of response in situations of distrust. In an analogous way to the counter-scenario generation that occurs when individuals prepare to encode messages they are suspicious of, the cognitive system habitually activates nondominant associations in contexts of distrust.

Experiment 2

Faces are unique stimuli in many ways (Zebrowitz, 1997). Therefore, it is instructive to examine whether our findings generalize to manipulations of trust that do not involve faces. The participants in Experiment 2 received sets of questions and answers provided by several individuals. In the trust condition, the participants were to use these answers to rate the spontaneity of the individuals. We assume that this orienting task presupposes trust. In the distrust condition, the experimental participants were told that the answers had been given by men or women and that although the women had been asked to provide truthful answers, the men had been asked to impersonate a woman in answering. Participants in the distrust condition were to determine whether the communicator was truthful (a woman) or deceptive (a man). We assume that this orienting task induced distrust and suspicion. Between reading the answer sets of different individuals, participants performed the priming task from Experiment 1 (of course, without the presence of faces). We explored whether the type of prime (congruent vs. incongruent) interacted with the orienting task (trust vs. distrust) analogously to Experiment 1.

Method

Participants

Forty-nine students at the Hebrew University participated for either course credit or payment. One student was eliminated because he was not a native Hebrew speaker.

Constructions of the Stimuli

We approached male and female students in the library and asked them to fill out a short questionnaire. Women were asked to answer truthfully. Men were asked to impersonate a woman. We informed the male students that their answers would be given to other students who would try to separate answers made by male imposters from those of female truth tellers. In the experiment proper, we used the answers given to five questions: “what do you like to do in your leisure time?,” “what do you think about plastic surgery?,” “explain in a few words how to change a tire on a car,” “describe what’s in your handbag right now,” and “describe what kind of gesture you would consider romantic.” The responses to each question varied in length from several words to several sentences. We chose five sets of replies (three men’s, two women’s) to be used as stimuli in the experiment proper.

Procedure

Respondents were informed that the experiment included two interwoven tasks. For practice, they were then shown the questions and answers given by one person. They were told that they could read and think about these answers for as long as they needed. They were also informed about the judgment dimension, but they were asked to withhold judgment until the end of the second task. After reading the questions and answers, the respondents were given a practice block (20 trials) of priming, as in Experiment 1. Immediately after the priming block, they were instructed to think again about the answers of the stimulus person and make a judgment.[2] The kind of judgments that respondents made is the experimental manipulation of trust/distrust. Specifically, in the trust condition respondents were to rate the spontaneity of that person and, in the distrust condition, to rate how confident they were that the person is a woman (truth teller) or an imposter man. Ratings were made on a 5-point scale (1 = *not spontaneous/definitely imposter*, 5 = *very spontaneous/certainly a female*).

This sequence (reading answers to five questions, 20 trials of priming task, making a judgment) was repeated four more times, with new sets of answers, and new prime–target pairs in each round. Thus, in addition to the practice session, the respondents rated four experimental persons (two men and two women) and had 80 trials of priming. The prime–target pairs were identical to those used in Experiment 1. However, unlike Experiment 1, the prime–target pairs were not presented together with faces.

Design

The experiment consisted of one between-subjects factor: the orienting task. Half the participants rated the spontaneity of each person who provided the answers; the remaining rated whether the person was an imposter (men) or a truth teller (women).

Results

We eliminated from the analyses the data of 1 respondent whose error rate in the priming task was above 20%. In 102 cases ($<$ 3% of the trials),[3] respondents failed to respond within the 2-s response window. These trials were eliminated from the analyses. Also, we eliminated response latencies from 67 trials that were more than 3 standard deviations from the mean (less then 2%). Finally, the response accuracy was examined. Overall, the accuracy was high (about 94%). In the analyses of the response latency, trials with incorrect responses were excluded. Thus, for each respondent we computed the mean latency of responding correctly to target words in the different face–prime–target combinations. The means are listed in Table 2.

A two-way mixed-model ANOVA, with trust/distrust as a between-subjects factor and type of prime as a within-subject factor, was performed on the latency for correct responses. The pattern of results was similar to that of Experiment 1. Specifically, the interaction contrast, comparing congruent versus incongruent primes in conditions of trust versus distrust, was significant, $F(1, 45) = 5.10$, $p < .05$, indicating that the facilitation of target adjectives by congruent and incongruent primes differed under conditions of trust and distrust. In contrast to Experiment 1, however, analysis of the simple effect within the trust and the distrust conditions failed to reject the null hypothesis of no difference between the facilitation of congruent and incongruent probes, $t(24) = 0.68$, for the distrust condition, $t(24) = -0.26$, for the trust condition. We believe that this failure reflects a loss of power in the simple-effect analyses because of the between-subjects design of the experiment. Specifically, let A designate the between-subjects factor (trust/distrust) and B the within-subject factor (congruent/incongruent primes). The error term for examining the simple effect of A within a level of the factor B is the mean of $MSE(A)$ and $MSE(B)$ (Winer, 1971, p. 545). It is thus heavily influenced by the variation between subjects. In our analysis, the between-subjects error term ($MSE[A]$) was over 10 times larger than the within-subject error term ($MSE[B]$), leading to a serious loss of statistical power in comparison with the statistical test of the 2 $\times$ 2 interaction contrast.[4]

Table 2 also shows that the respondents' accuracy was high. A two-way ANOVA also indicates that accuracy in noun trials was higher than that in adjective trials, $F(1, 45) = 10.05$, $p < .05$. However, accuracy was not different in the congruent and incongruent trials, $F(1, 45) = 1.45$, $p > .2$, nor was there a significant interaction ($F < 1$). Thus, as in Experiment 1, there is no indication that the pattern of latency results reflects a speed–accuracy trade-off.

Experiment 3

Experiments 1 and 2 both used the extent of priming facilitation as a marker for the difference in the kind of associations activated by the prime word under trust and distrust. Experiment 3 not only examines whether the effect is specific to this measure; more important, it tests our hypothesis more directly by comparing the actual frequency of congruent and incongruent associations activated under conditions of trust and distrust. Participants in the experiment were shown trustworthy or untrustworthy faces, as in Experiment 1. In contrast to Experiment 1, however, each face was accompanied by a single word. Participants were asked to generate an association to each word. Later, participants were probed for their memory of the words presented with the faces. According to our theoretical analysis, when individuals think about a word presented in the context of distrusted faces, they are more likely to spontaneously activate concepts that are *in*congruent with that word. If so, this should have two related effects, one on the kind of free associations generated in response to the target words and the other on memory for the target words. Specifically, we hypothesize, first, that untrustworthy faces lead to a greater tendency to generate incongruent associations than do trustworthy faces. Second, memory for the target words ought be less crystallized following distrust. In particular, because we hypothesize that respondents think about incongruent associations at the time they encode the target words under distrust, we expect two kinds of memory mistakes: (a) *new* foils incongruent with the *old* targets should be mistaken for *old* words to a greater extent under distrust than under trust and (b) *old* target words should be more likely to be mistaken for *new* ones under distrust. Thus, for example, if a respondent activates *constant* as an association of *transient* during

---

[2] We believe that this procedure increases the chance that the trust or distrust elicited by the orienting task remains during priming.

[3] This represents an increase relative to Experiment 1. We believe that it has to do with the cognitive load of the respondents, who had to keep in memory the answers from the spontaneity/male–female task.

[4] Although it may seem that there is a large difference between the speed of processing of nouns that appear following trustworthy faces ($M = 786$) and those following untrustworthy faces ($M = 811$), the null hypothesis of no systematic difference could not be rejected, $F(1, 45) = 0.56$, $p > .3$.

Table 2
Means (and Standard Deviations) of the Latency of Accurate Responses (in Milliseconds) and Proportion of Correct Responses From the Adjective–Noun Classification Task (Experiment 2)

Rating task	Congruent prime adjective target	Incongruent prime adjective target	Irrelevant prime noun target
Distrust (M/F)	851 (138)/.92	840 (133)/.94	811 (115)/.96
Trust (spontaneity)	825 (123)/.94	850 (137)/.95	786 (116)/.98

Note. M/F = male or female.

the association-generation phase, he or she should be more likely to mistakenly treat the word *permanent* (i.e., a new incongruent foil) as an *old* target and less likely to correctly remember *transient.*

Method

Participants

Forty students at the Hebrew University participated for either course credit or payment. They were allocated randomly to one of the two experimental conditions.

Procedure

The experiment consisted of several parts. In Part 1, participants practiced the free-association task. In each trial they were presented with a word on the screen. They were asked to generate another word as a free association. They were given a time-limited window of response for the generation. Specifically, they had a 3-s window for the first 10 trials and a progressively shorter window for the remaining 30 trials. In the last 10 trials the respondents were allowed only a 2-s interval to generate the free association. This progressive shortening was used to familiarize the participants with the free-association task and to train them to respond quickly, thus reducing the likelihood that they could control the associations they generate. Responses (i.e., free associations) were recorded on an audiotape. There were 40 trials in this part.

In the second part of the experiment, the participants underwent the induction phase, which was identical in procedure to Experiment 1. On each trial of the third part of the experiment the participants saw a face. After 800 ms, a word was superimposed on the face just below the eyes. They then had to generate a free association to the word. Generation was constrained by a 2-s response window. A bell and the disappearance of the word (as well as the face) signaled the end of the response window. As in Part 1, responses were recorded. There were 42 trials in this part of the experiment. The first 6 trials were used as a warm-up, and included filler words. The remaining 36 trials included target-adjective words from the list used in Experiment 1.

Part 4 of the experiment served as a 5-min buffer. Respondents were given several series of numbers. For each series they had to figure out the rule that underlay the series elements and write down the next element. During the next and final part of the experiment the participants' memory was probed. On each trial they were presented with a single word and were asked to indicate whether it had appeared during generation (i.e., Part 3). This was done on an 8-point scale (1 = *certainly new* to 8 = *definitely appeared*). There were 57 trials in this part. The first 3 memory probes were selected from the filler words in Part 3. The remaining words consisted of three sets: 18 words were old (i.e., words that were targets for generation in Part 3), 18 words were the incongruent associations of words from Part 3, and 18 words were completely new, namely, words that were unrelated to Part 3 words.

Counterbalancing

The words in Part 1 were different from the words in Parts 3 and 5. The pictures in Part 2 (induction) were different from those in Part 3 (free-association task). The 36 target words in Part 3 consisted of two 18-word lists. In the memory part, half of the participants saw one list as the old words and the incongruent associations of words from the other list. The remaining participants saw the other list as old and the incongruent associations of words from the first list. The order of presentation of the target words for generation and the order of the memory probes was random.

Design

The experiment consisted of one between-subjects factor: the type of face that accompanied the target words during free association. Thus, unlike in Experiment 1, following the induction, the respondents saw either trustworthy or untrustworthy faces.

Results

Two individuals who were blind to the experimental condition coded the participants' associations. Because of mechanical failure, the recordings of 5 participants could not be transcribed. For these persons, therefore, we have only the memory results. The coders classified the generated words according to whether they were incongruent with the stimulus word presented on the screen. The coders agreed in 96% of the cases. Disagreements were resolved by a third person. The incongruent associations included responses such as *dirty* (for *clean*), *passive* (for *active*), or *disgusting* (for *tasty*). The nonincongruent responses varied in the type of stimulus–response relationships. Some responses had meanings similar to the stimulus word (e.g., *neat* [for *clean*], *hard* [for *rigid*]), others were linked semantically to the stimulus word (e.g., *money* [for *spendthrift*], *palace* [for *magnificent*], or *knife* [for *sharp*]), and still other responses were linked to the stimulus word through a familiar expression (e.g., *partly* [for *cloudy*], *bird* [for *rare*]). In less than 10% of the cases the coders could not identify the associative link between the stimulus word and the response.

Table 3 presents the means of the total number of words and the number of incongruent words that were generated. First, note that respondents were successful in generating a word within the 2-s response window in about 86% of the cases, averaging around 31 words (out of the 36 trials). The total number of words did not

Table 3
Means (and Standard Deviations) of the Number of Words Generated During the Free-Association Task (Experiment 3)

Face	Total no. of words	No. of incongruent words
Untrustworthy	31.68 (2.47)	11.87 (8.73)
Trustworthy	30.72 (4.47)	7.38 (6.05)

differ significantly as a function of the presence of trustworthy or untrustworthy faces, 31.68 versus 30.72, $t(27) = 0.77$, $p > .4$.[5] Our analysis predicted that the level of (dis)trust associated with the faces would influence the generation of incongruent associations. As Table 3 shows, this was indeed the case. Participants tended to generate more incongruent associations under distrust than under trust, 11.87 versus 7.38, $t(26) = 1.72$, $p < .05$, one-tailed.

In the last part of Experiment 3, respondents were presented with three types of memory probes: *old* (words that had been targets for generation), *new incongruent associations* (words that were incongruent with targets for generation), and *completely new* (words that were unrelated to the targets). Respondents indicated their recollection of whether each probe word had been displayed during the generation part (Part 3) using an 8-point rating scale (1 = *certainly new* to 8 = *definitely appeared*). Table 4 presents the mean memory ratings of the three types of memory probes. Our analysis predicted that, because of the activation of incongruent associations during the free-association task, the representations of the target words would be less crystallized for the participants who encoded these words in the context of distrusted faces. That is to say, when words were encoded in the context of distrust, the two classes of words, original old words and their incongruent associations, would be more confused in memory.

As Table 4 shows, compared with participants who generated free associations in the presence of trustworthy faces, those who generated associations in the presence of untrustworthy faces rated the old memory probes lower and the new incongruent probes higher. A two-way mixed-model ANOVA (Faces × Type of Memory Probe) indicated, not surprisingly, that the three types of probes were rated very differently, $F(2, 76) = 348.88$, $p < .01$. Our analysis was mute with respect to the effect of trust/distrust on the memory of the new unrelated probes. We did correctly predict, however, that respondents would confuse more of the old probes and the new incongruent probes under distrust than under trust. This theoretical prediction was tested by a 2 (type of face) × 2 (old vs. new incongruent) interaction contrast, which was significant, $F(1, 38) = 6.74$, $p < .05$. As hypothesized, respondents in the distrust condition tended to perceive the old probes as less old and the new incongruent probes as less new, suggesting that during the free-association task respondents accessed incongruent associations. Simple-effect analyses revealed that the differences between trust and distrust conditions were statistically significant for old probes, $t(38) = 3.02$, $p < .01$; marginally significant (though in the opposite direction) for new incongruent probes, $t(38) = -1.53$, $p < .07$, one-tailed; and nonsignificant for new unrelated probes, $t(38) = -.37$, $p > .5$.

Table 4
Means (and Standard Deviations) of Memory Ratings as a Function of the Face Condition and Type of Memory Probe (Experiment 3)

Face	Old	New incongruent	New unrelated
Untrustworthy	6.42 (0.72)	3.54 (1.13)	2.07 (0.60)
Trustworthy	7.04 (0.58)	2.96 (1.25)	1.99 (0.76)

Note. Ratings were made on an 8-point scale (1 = *completely new,* 8 = *definitely appeared*).

A second index of crystallization is based on signal detection theory. This procedure allows us to determine whether the trustworthiness of the faces influenced the respondents' ability to discriminate old from new recognition probes. Our prediction, of course, is that discrimination sensitivity *de*creases when the target words are encoded in the presence of distrusted faces (see Schul et al., 1996). For each respondent, we computed a bias-free sensitivity index by comparing decisions about the incongruent and the old probes. This was done by considering ratings of 1–4 as indicating new and ratings of 5–8 as indicating old. For each respondent, we computed the sensitivity index A' in the procedure outlined by Grier (1971). Analysis of this index revealed that memory discriminations of respondents who responded to the target words in the presence of untrustworthy faces were less sensitive than those of respondents who saw the words in the presence of trustworthy faces, .81 versus .86, $F(1, 39) = 4.26$, $p < .05$.

A final analysis compared the correlations between the indices of association generation and memory. The results revealed that although the correlations were in the expected direction, they were small. The number of incongruent words generated during the free-association task was negatively correlated with the rating of old memory probes and positively correlated with the ratings of incongruent probes (rs = −.07 and .13, *ns*). Obversely, the number of words that were not incongruent was unrelated to the memory of the old probes and negatively correlated with the memory of the incongruent probes (rs = –.01 and –.27, *ns*). These findings suggest that, although the generation and memory measures are influenced by a common factor (i.e., the nature of faces), a variety of different factors also affect them, and as a result, they are only weakly related.

General Discussion

Past research has shown that in preparing to cope with a potentially invalid message, receivers increase the complexity of their processing by trying to construct alternative scenarios, one in which the message is valid and another in which it is not. Such a strategy is useful in facilitating successful discounting of invalid information, combating the correspondence bias, and reducing (message) order effects. The current series of studies examines receivers' propensity to activate message-incongruent cognitions when (a) their distrust is unrelated to the message and (b) they cannot control the activation intentionally.

Participants in Experiment 1 were presented in each trial with a face that elicited either trust or distrust. Shortly afterward, they saw two words, a prime and a target, and had to indicate whether the target was an adjective or a noun. The results show that the pattern of priming facilitation on processing of the target words depended on the type of the face. When the face was associated with trust, the congruent primes facilitated the processing of the target adjectives more than the incongruent primes did. However, when the face was associated with distrust, the opposite occurred;

[5] As the variances of the two groups were not equal, degrees of freedom were adjusted on the basis of the formula suggested by Satterthwaite.

namely, the incongruent primes facilitated processing of the targets more than the congruent primes. Experiment 2 replicated this pattern using a very different manipulation of trust/distrust. In the trust condition, respondents evaluated the spontaneity of a protagonist on the basis of a series of answers she gave. In the distrust condition, respondents evaluated whether the protagonist was truthful or whether he was trying to masquerade as another person. All respondents were given a priming task, similar to that in Experiment 1. As in Experiment 1, the pattern of facilitation associated with the two types of primes was different under trust than under distrust.

Experiment 3 assessed the activation of congruent and incongruent associations more directly. As in Experiment 1, respondents were presented with faces that elicited either trust or distrust. They were also given a target word and asked to provide an association for it. Later, their memory for the target words was probed. It was found that respondents generated more incongruent associations in the presence of distrusted faces than in the presence of trusted faces. Moreover, respondents' memory for target words encoded in the presence of distrusted faces was less distinctive than their memory for target words encoded in the presence of trusted faces. In particular, under distrust the old target words were considered less likely to have been shown, whereas new foils that were opposite of the original target words were considered as more likely to have been shown. This pattern is consistent with the suggestion that under distrust, more than under trust, respondents spontaneously thought about concepts incongruent with the primes.

We believe that these findings are consistent with the suggestion that the trust/distrust distinction provides a basic context for information processing (Cosmides & Toby, 1992). When they believe a source, receivers tend to concentrate on message-congruent associations. However, if the source is suspected of being untrustworthy, receivers spontaneously activate message-incongruent associations, as they are considering what might happen if the message is invalid.

In many respects, receivers who distrust a message do something similar to individuals who muse about "what might have been," that is, engage in counterfactual thinking (e.g., Roese, 1997). Galinsky and Moskowitz (2000) recently reported that when the mode of counterfactual thinking has been activated, respondents are more likely to come up with a new use of an object or to test the reverse of a given hypothesis, and they are less likely to show a confirmation bias. Thus, like receivers who suspect that a message is invalid, individuals who are induced to consider counterfactual ideas tend to activate stimulus-incongruent associations. Our findings complement this research by showing that the activation of distrust has effects similar to counterfactual thinking even though the procedure (i.e., thinking about the alternative) is not activated directly.

We think that the activation effects observed in the present study reflect a well-practiced generalization of an information-processing strategy people use to cope with potentially deceptive messages. If people spontaneously generate counterinferences when processing a suspect message, senders of communications should be sensitive to the danger of having a reputation as an untrustworthy source. Indeed, reputation plays a special role in interpersonal interactions among humans as well as many primates (Andrews, 2002; Bacharach & Gambeta, 2001; Kollock, 1994; Kramer, 1999). Cheney and Seyfarth (1991) observed that when the calls of a particular monkey are unreliable, others in the troop ignore them, as if the caller had a reputation for unreliability. Moreover, a reputation of this sort was formed only when the unreliability was consistent, namely, when the monkey's signals were unreliable regardless of the referent domain. Humans behave differently. We often attribute unreliability to an intention to deceive; not only do we infer this intention from a single instance of being misled, but once the inference is made, we loath to be persuaded otherwise. This contrasts sharply with a reputation for reliability or honesty, which requires a good deal of evidence and frequent reinforcement to be maintained (Gidron, Koehler, & Tversky, 1993). Taken together, this implies that human senders should be particularly cautious about gaining a reputation as untrustworthy, as receivers may interpret communication that comes from an untrustworthy source contrary to the intended meaning (cf. Mayo, Schul, & Burnstein, in press).

Although distrust might suggest a challenging and suspenseful relationship and trust a relationship of boring predictability, it is more often the case that trust is experienced as a pleasant state of affairs and distrust as an unpleasant one (Hardin, 2001; McCornack & Parks, 1986; Messick & Kramer, 2001). This raises the possibility that the trust/distrust effects in our study were due to an emotional reaction of this sort rather than to suspicion per se. Accordingly, one might argue that it was the negative affect elicited by the distrusted faces (Experiments 1 and 3) or by the effort needed to cope with potential deception (Experiment 2) that explains our results. There are at least two reasons that this interpretation is unlikely, in our opinion. First, Experiment 1 manipulated trust/distrust within an experimental session, so that respondents would have had to oscillate between a positive and negative affective state several times in a minute, making for considerable confusion and rendering either state less significant. Second, past research has shown that negative moods can lead to greater reliance on stimulus properties, whereas positive moods can lead to more top-down processing (see review in Fiedler, 2001). It seems difficult to make sense of our results within this framework. On the one hand, it could be argued that accessing the (expected) congruent association is stimulus-based processing, but this would be inconsistent with the findings on the influence of a positive mood. To be sure, there is also the alternative argument, namely, that accessing incongruent associations reflects stimulus-based thinking. This might be consistent with the effect of negative mood states, and we are currently comparing the two interpretations more directly.

At present, our results seem best interpreted as reflecting a low-level general mechanism that allows people to cope with uncertainties about what someone else is telling them. When people trust the message, they access its dominant meaning—they believe the message. However, when there is distrust, this mechanism causes the receiver to suspend belief—for the moment the message is not accepted. Past studies of strategic thinking under conditions of distrust have observed an increase in the complexity of encoding of the message information, as if receivers are trying to protect themselves from being misled by testing potential counter-scenarios (Fein, 1996; Fein et al., 1997; Schul, 1993; Schul et al., 1996). The present research suggests that (a) even when the distrust is unrelated in any meaningful way to the message and (b) even when receivers are unable to prepare a strategic response, the cognitive system reacts to distrust by automatically inducing the consideration of incongruent associations—it seems designed to ask, "and what if the information were false?"

References

Akehurst, L., Koehnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology, 10,* 461–471.

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom. *Journal of Nonverbal Behavior, 23,* 66–89.

Andrews, P. W. (2002). The influence of postreliance detection on the deceptive efficacy of dishonest signals of intent: Understanding facial clues to deceit as the outcome of signaling tradeoffs. *Evolution and Human Behavior, 23,* 103–121.

Bacharach, M., & Gambeta, D. (2001). Trust in signs. In K. S. Cook (Ed.), *Trust in society* (pp. 148–184). New York: Russell Sage Foundation.

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

Berry, D. S., & Waro-Finch, J. L. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of Personality and Social Psychology, 61,* 497–520.

Cheney, D. L., & Seyfarth, R. M. (1991). Truth and deception in animal communication. In C. A. Ristau (Ed.), *Cognitive ethology: The minds of other animals* (pp. 127–151). Hillsdale, NJ: Erlbaum.

Cosmides, L., & Toby, J. (1992). Cognitive adaptation for social exchange. In J. E. Barkow, L. Cosmides, & J. Toby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

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

DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. T. Gilbert, S. T. Fiske, & G. Lindsey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 3–40). Boston: McGraw-Hill.

DePaulo, B. M., & Kashy, D. H. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74,* 63–79.

DePaulo, B. M., Kashy, D. H., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979–993.

Dyer, R. F., & Kuehl, P. G. (1978). A longitudinal study of corrective advertising. *Journal of Marketing Research, 15,* 39–48.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46,* 913–920.

Faces 3.0 [Computer software]. (1998). Quebec, Canada: InterQuest.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology, 70,* 1164–1184.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and correspondence bias. *Journal of Personality and Social Psychology, 58,* 753–764.

Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial and inadmissible testimony. *Personality and Social Psychology Bulletin, 23,* 1215–1226.

Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology, 24,* 163–170.

Festinger, L., & Maccoby, N. (1964). On resistance to persuasive communications. *Journal of Abnormal and Social Psychology, 68,* 359–366.

Fiedler, K. (2001). Affective influences on social information processing. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 163–185). Mahwah, NJ: Erlbaum.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652–671.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology, 36,* 384–409.

Gidron, D., Koehler, D. J., & Tversky, A. (1993). Implicit quantification of personality traits. *Personality and Social Psychology Bulletin, 19,* 594–604.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733–740.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65,* 221–233.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Griffin, D. R. (1992). *Animal mind.* Chicago: The University of Chicago Press.

Hardin, R. (2001). Conceptions and explanations of trust. In K. S. Cook (Ed.), *Trust in society* (Vol. 2, pp. 3–39). New York: Russell Sage Foundation.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78,* 837–852.

Hill, T., Lewicki, P., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology, 26,* 350–371.

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin, 19,* 501–512.

Inquisit 1.32 [Computer software]. (2002). Seattle, WA: Millisecond Software.

Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology, 100,* 313–345.

Kramer, R. M. (1999). Trust and distrust in organization: Emerging perspectives, enduring questions. *Annual Review in Psychology, 50,* 60–98.

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge.* New York: Plenum Press.

Kruglanski, A., & Freund, T. (1983). The freezing of lay inferences effects of impressional primacy, ethnic stereotyping and numerical anchoring. *Journal of Experimental Social Psychology, 19,* 448–468.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review, 23,* 438–458.

Mayo, R., Schul, Y., & Burnstein, E. (in press). "I am not guilty" versus "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology.*

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Communication Yearbook, 9,* 377–389.

McGraw, K. M., Lodge, M., & Jones, J. M. (2002). The pandering politicians of suspicious minds. *Journal of Politics, 64,* 362–383.

Messick, D. M., & Kramer, R. M. (2001). Trust as a form of shallow morality. In K. S. Cook (Ed.), *Trust in society* (pp. 89–117). New York: Russell Sage Foundation.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory:

Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Newman, L. S., & Uleman, J. S. (1989). Spontaneous trait inference. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 155–188). New York: Guilford Press.

Petty, R. E., Wells, G. T., & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34,* 874–884.

Pew Research Center for the People and the Press. (1999). *Striking the balance, audience interests, business pressures and journalists' values.* Retrieved February 18, 2004, from http://people-press.org/reports/display.php3?ReportID=67.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121,* 133–148.

Schul, Y. (1993). When warning succeeds: The effect of warning on success of ignoring invalid information. *Journal of Experimental Social Psychology, 29,* 42–62.

Schul, Y., & Burnstein, E. (1998). Suspicion and discounting: Ignoring invalid information in an uncertain environment. In J. M. Golding & C. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 321–348). Mahwah, NJ: Erlbaum.

Schul, Y., Burnstein, E., & Bardi, A. (1996). Dealing with deceptions that are difficult to detect: Encoding and judgment as a function of preparing to receive invalid information. *Journal of Experimental Social Psychology, 32,* 228–253.

Snyder, M. L., & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology, 14,* 148–162.

Tellis, G. J. (1997). *Advertising and sales promotion strategy.* New York: Addison-Wesley.

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

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Oxford, England: Harvard University Press.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 446–496). Boston: McGraw Hill.

Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49,* 338–346.

Wegner, D. M., Wenzlaff, R., Kerker, R. M., & Beattie, A. M. (1981). Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology, 40,* 822–832.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

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

Received June 3, 2003
Revision received January 8, 2004
Accepted January 13, 2004 ■