Thinking the opposite under distrust: Do untrustworthy faces facilitate the recognition of antonyms?

Corina T. Ulshöfer[1*], Nicole Ruffieux[1] and Margit E. Oswald[1]

[1]Institute for Psychology, Social Psychology, University of Bern, Switzerland; E-mail: corina.ulshoefer@psy.unibe.ch*, nicole.ruffieux@psy.unibe.ch, margit.oswald@psy.unibe.ch

[*]Corresponding author at: Institute for Psychology, Social Psychology, Fabrikstrasse 8, 3012 Bern, Switzerland. E-mail: corina.ulshoefer@psy.unibe.ch (C. T. Ulshöfer); Phone: 0041/31/6314028; Fax: 0041/31/6318212*

* New contact: corina.ulshoefer@erz.be.ch

Abstract

Distrust should automatically activate a "thinking the opposite". Thus, according to Schul, Mayo and Burnstein (2004), participants detect antonyms faster when confronted with untrustworthy rather than trustworthy faces and detect synonyms faster when confronted with trustworthy rather than untrustworthy faces. In an induction phase the authors tried to amplify the (un-)trustworthiness of faces by combining them either with true or false statements. In three experiments we tested whether the interaction effect is due to the induction phase, and whether the canonicity of antonyms plays a crucial role. Additionally, we adapted faces more according to European standards and presented stimuli more with regard to priming rules. Results show that the interaction effect cannot be reliably predicted. Even if faces are culturally adapted and priming rules are applied more strictly, an interaction effect depends on whether the induction phase is applied and on the canonicity of antonyms rather than on the trustworthiness of faces. Generally, our results show the inappropriateness of a paradigm that was supposed to test a thinking the opposite under distrust.

(172 words)

Keywords: distrust, thinking the opposite, information processing

Thinking the opposite under distrust:

Do untrustworthy faces always facilitate the recognition of antonyms?

Being in a state of distrust may sometimes improve problem solving and judgmental veracity. Thus, researchers have found that people under distrust are better able to think in multiple categories (Friesen & Sinclair, 2011), create counter-scenarios (Schul, 1993; Schul, Burnstein, & Bardi, 1996), avoid a correspondence bias (Fein, 1996), perform better on reasoning tasks (Gigerenzer & Hug, 1992), be more creative (Mayer & Mussweiler, 2011), and think the opposite more easily (Schul et al., 2004).

The creation of counter-scenarios (Schul et al., 1996) and the search for non-routine contingencies (Schul, Mayo, & Burnstein, 2008) can be considered as deliberate processes or at least reflect conscious strategies that have probably been automated. Alternatively, a “thinking the opposite” can be seen as an unconditioned, unconscious reaction, as documented in the frequently cited Experiment 1 of Schul et al. (2004). The results of this experiment are intriguing: when participants were primed with an adjective superimposed on a face presented on a screen, they recognized synonyms faster than antonyms if the face was trustworthy. However, antonyms were recognized faster than synonyms if the face was untrustworthy. Nevertheless, questions arise whether the interaction effect will reliably occur without additionally enhancing the trustworthiness of faces by an induction phase, and when the canonicity between adjectives and their antonyms has been controlled.

Having changed the conditions accordingly, we critically tested the appropriateness of the paradigm that was invented by Schul et al. (2004) and examined whether their interaction effect can be replicated. Our studies give rise to the assumption that the effect is due to the induction phase, and is importantly influenced by whether the opposite of an adjective is unique (canonical antonym) or ambiguous (non-canonical antonym).

Deliberate Information Processing under Distrust

In daily life, people are in general ready to trust the statements of others and to consider them as true. This well-developed "truth bias" may be considered as an heuristic that not only simplifies the cognitive processes during the interaction but facilitates communication and contributes to the maintenance of social relations (Stiff, Kim, & Ramesh, 1992).

Of course, there may be information or situational contexts that raise distrust. Interpersonal distrust may arise because of negative experiences with the other person (Lewicki & Bunker, 1996), but also because of social stereotypes, like specific features of a face (Todorov, Baron, & Oosterhof, 2008). Many authors assume that distrust not only reduces the truth bias but also results in more deliberate processing of information (e.g. Stiff et al., 1992). The state of distrust indicates danger, that is, the possibility that your interaction partner may want to take advantage over you. While trust means renouncing social control totally or at least partially (Mayer, Davis, & Schoorman, 1995), in the state of distrust it appears necessary to assess the intentions and motives of others more closely. More cognitive activity is required (Gilbert, 1993).

In the case of attributed ulterior motives, Fein and colleagues (Fein, McCloskey, & Tomlinson, 1997) have shown that participants process information as if they were analyzing the distrusted persons" actions from two different points of view: one consistent with the explicitly claimed motive, the other consistent with the ulterior motive. When individuals are suspecting the motivation or the intention of a source, they either apply non-routine solution processes (Schul et al., 2008) or attempt to resist persuasion by counterarguing (Schul, 1993). In preparing to receive invalid information, people may exert themselves to build counter-scenarios around particular pieces of information that they believe to be invalid (Schul, 1993; Schul et al., 1996).

Automatic Information Processing under Distrust

While the creation of counter-scenarios presupposes a more or less deliberate processing of the given information, it is of interest whether people in a state of distrust even automatically activate ideas incongruent with those in the message (Schul et al., 2004). The automatically activated thinking the opposite, however, does not imply that the given information is analyzed from different perspectives but rather that a kind of rejection happens which activates the contrary meaning: for example, the automatic and faster activation of "cold" if the person says "warm".

Schul et al. (2004) showed that cues of untrustworthiness, even irrelevant to the information to be rated, can positively influence consideration of incongruent associations. In Experiment 1, in the induction phase (Phase 1), they generated an association between a certain facial expression and the truth of a statement. Faces with narrow eyes (cue for untrustworthiness; Zebrowitz, 1997) were associated with false statements (A), whereas faces with round eyes (cue for trustworthiness) were associated with true statements (B). In the priming phase (Phase 2), faces from category A or B were presented together with adjectives which were followed by another related adjective or an unrelated noun. The participants were to indicate as rapidly as possible whether the second word was an adjective or a noun. The pairs of adjectives were either *semantically congruent (synonyms,* e.g., *cold-cool*) *or incongruent (antonyms, e.g. cold-warm)*. The results of this experiment showed no main effect but demonstrated a significant interaction effect (see Appendix A for effect size and power analyses). In the trust condition, response latencies were shorter only for synonyms. In the distrust condition, by contrast, response latencies were shorter for antonyms. According to Schul et al. (2004), the interaction effect corroborates the assumption that an activation of incongruent associations is a generalized pattern of response in a state of distrust. The results of this experiment are fascinating. However, two issues are threatening their generalizability. It is unclear whether the additional induction phase itself and not the priming phase with the trustworthy and untrustworthy faces may have produced the effects. Furthermore, adjectives can have unique antonyms (canonical) or ambiguous antonyms (non-canonical; Paradis, Willners, & Jones, 2009). With a non-canoncial antonym a thinking the opposite of an adjective may not necessarily lead participants‟ association to the same antonym as the one presented in the experiment. Thus, it remains an open question whether untrustworthy faces will lead to a faster detection of antonyms independent of their canonicity.

Empirical Predictions

Induction phase

The induction phase of the experiment may be a confounding variable. In this phase, an association between a facial expression and the truth of a statement is created: participants learn that a trustworthy face is always accompanied by a true statement, and that an untrustworthy face is always accompanied with a false statement. This additional induction phase was applied by Schul et al. (2004) to ensure that faces with narrow eyes are distrusted and faces with round eyes are trusted. However, the induction itself may have probably caused the experimental effect, independent of the trustworthiness of faces. It is unclear whether it has been the faces that primed distrust or rather the true and false statements in the induction phase, which were systematically paired with trustworthy vs. untrustworthy faces. In other words, it is unclear whether the (un)trustworthiness of faces is an unconditioned stimulus, as Schul et al. (2004) imply, or rather a conditioned stimulus due to the classical conditioning with true vs. false statements.Weil (2010) has shown that even geometrical figures, which she used instead of faces, work as a prime for truth in a follow-up experiment.Therefore, a replication of the priming paradigm, with and without the induction phase, may provide additional insight into the cognitive processes that cause the interaction effect. According to Schul et al. (2004), the induction phase should only enhance the genuine effect of trustworthy and untrustworthy faces. One can expect the same interaction effect under both conditions, yet the effect can possibly be more pronounced if an additional induction phase is applied. However, if we were to find no such interaction or even an opposite interaction effect without applying the induction phase, doubts would arise that untrustworthy faces always enhance the recognition of antonyms.

Canonicity

The experiment of Schul et al. (2004) does not take into consideration the distinction between canonical and non-canonical antonyms (Paradis et al., 2009). A canonical antonym, e.g. “short-long”, may be processed much faster than a non-canonical, ambiguous antonym such as “easy-difficult”. In this case, one may also view “complicated” or “challenging” as proper antonyms of “easy”. Thus, whenever canonical antonyms are used as prime-target pairs, one should expect faster response latencies in comparison with non-canonical antonyms. In general, canonical antonyms should be processed faster than synonyms and synonyms should be processed faster than non-canonical antonyms (Paradis et al., 2009).

More interestingly, however, is the question whether even non-canonical antonyms are generally detected faster under distrust. At least, sometimes participants in a state of distrust may even react more slowly when recognising non-canonical antonyms. Take, for example, a case when “easy” is presented as a prime, and a participant associates not the target word “difficult” but another antonym such as “complicated”. This participant may be slower because he has to correct his first association, his first opposite, and react on another opposite which was not activated in his mind. Alternatively, a participant under trust is not supposed to think the opposite and has therefore not to make this detour. We assume that a main effect will appear when only canonical antonyms are used. For non-canonical antonyms, no precise hypothesis can be postulated due to the described ambiguity of these word pairs.

Minor issues

One further issue refers to the ethnicity of faces. The faces used by Schul et al. (2004) may represent more the typical ethnical diversity in the Near East, and these faces have either round or narrow eyes. Although round eyes are a general cue for trustworthiness and narrow eyes usually stand for untrustworthiness (Gorn, Jiang, & Johar, 2008), face perception is culturally dependent. Perception of faces from foreign cultures is, for example, less detailed than of faces of ones" own culture (Papesh & Goldinger, 2010). Thus, it can be questioned whether Central Europeans distinguish between the original trustworthy and untrustworthy faces as well as participants from Israel. However, even if we expect a smaller influence of the original faces, we would expect to replicate the results of Schul et al. (2004) as soon as we use trustworthy and untrustworthy European faces.

The other issue refers to priming rules. When analyzing the experiment under a priming perspective, the SOA of only 82 ms is unusually short for supraliminal priming experiments (Wentura & Degner, 2010) [1]. Especially, the Inter Stimulus Interval (ISI) of 0 ms is probably too short to guarantee the conscious perception of the prime and besides this short period of time may produce a masking effect of the prime due to the immediately following target. Furthermore, displaying a fixation cross before each trial will provide a better guarantee that participants focus on the prime. Finally, 80 trials are a very small number of trials for a priming experiment. We hypothesize that we will find stronger interactions with and without an induction phase, when we add ISI and when we display a fixation cross prior to each trial and increase the number of trials.

The Present Research

In the first of three experiments, we replicated Experiment 1 of Schul et al. (2004) with the original faces and pairs of adjectives, translated from Hebrew to German (Experiment 1). In Experiment 2, we created faces with Central European features and controlled them for their trust- and untrustworthiness. We also enlarged SOA and the number of trials, and in Experiment 3, we additionally manipulated the canonicity of antonyms. The application of the original induction was treated as an independent variable in all three studies. We report how we determined our sample sizes (Appendix A), all data exclusions (Appendix C), all manipulations, and all measures in the studies.

Experiment 1

Method

Design and participants. Experiment 1 was a replication of the experiment of Schul et al. (2004), supplemented with the variable "Induction". We employed a 2 (Target: Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) x 2 (Induction: Original Induction vs. No Induction) mixed design, using Target and Face as within-subjects factors. Furthermore, two control conditions "No Face" and "No Prime" from the original experiment by Schul et al. (2004) were adopted. In condition "No Face", polygons were used instead of faces in Phase 2, and in condition "No Prime", the prime word in Phase 2 consisted of a letter sequence without meaning. Participants were 103 students at the University of Bern, Switzerland, of which 94 participants were psychology undergraduates and 87 participants were women. The mean age of participants was 22.46 years (*SD* = 5.12). The participants were randomly assigned to one of four between-subjects conditions (Original Induction: 28; No Induction: 27; No Face: 24; No Prime: 24). An a prioripower analysis showed that with 24 participants in each condition, a power of 80% is achieved (see Appendix A for power analyses). They received course credit for their participation or a candy bar.

Materials and procedure. The same material was used as in the experiment of Schul et al. (2004): 40 faces with narrow eyes and 40 faces with round eyes. 20 faces of each type were used for the induction phase and 20 were used for the priming phase (Figure 1).

In a pretest, 46 participants rated to what extent the 40 original faces were untrustworthy, trustworthy, happy, sad, and neutral on 9-point Likert scales (see Appendix B, Figure B.1), one group with induction (N= 22) and one without (N=24). In general, untrustworthy faces were rated as being more untrustworthy than trustworthy and trustworthy faces were rated as being more trustworthy than untrustworthy. With induction, untrustworthy faces were rated less untrustworthy than without an induction, $t(44)= -1.31$, *MSE*= 1.54, *p*=.20, d= .39. For trustworthy faces, participants rated them more trustworthy in the condition with induction than without an induction, *t*(44)= 1.99, *MSE*= .88, *p*=.053, d= .59. Furthermore, trustworthy faces were rated quite high on "sadness" (Induction: *M*=4.24, *SD*=0.80; No Induction: *M*=4.56, *SD*=1.16).

Additionally, 20 correct and 20 false trivia sentences were created with low and high level of difficulty (low level of difficulty and correct: Buddhism is one of the five world religions.; high level of difficulty and correct: A genealogist is concerned with family trees.; low level of difficulty and false: Isaac Newton was a painter. high level of difficulty and false: Spanish painter Salvador Dali was an impressionist.). We used 20 pairs of synonyms and 20 pairs of antonyms and 40 non-related adjective-noun pairs. The experiment was programmed and run with Inquisit 3.0.6.0 (Milliseconds, 2012).

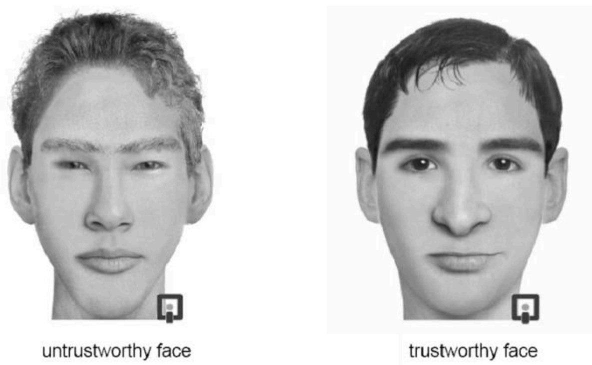


Figure 1: Example of original faces from Experiment 1 of Schul et al. (2004)

Entering the laboratory, participants signed an informed consent and sat down in front of a monitor[2]. Up to four participants took part in each session. In Phase 1, the induction, each of 20 narrow-eyed faces was presented with a false statement and each of 20 round-eyed faces was presented with a correct statement. Each face with each statement remained on the screen until the participant decided whether the statement was true or false. Participants were also asked to remember which face was presented with which statement. In all conditions, except in condition „No Induction", participants completed 80 trials of this task[3]. In Phase 2, the priming phase, either a narrow-eyed face or a round-eyed face was shown for 800 ms. Then a prime word was presented for 82 ms a little below the nose of the face. This prime word was immediately (ISI = 0 ms) replaced by a target word, which remained until the participant decided whether the target word was a noun or an adjective. If it was a noun, participants were indicated to press Key “S”, labeled with a blue sticker. If it was an adjective, they were indicated to press Key “K”, labeled with a yellow sticker. Participants had to make each of the 80 decisions within 2000 ms. Thereafter, they were thanked for their participation, and fully debriefed.

Dependent variables. Response latencies of each trial were measured and served as dependent variable. At the end of the experiment, participants had to specify their gender, age and field of study.

Eliminations. In total, 4 participants had to be excluded from the analyses (3 in “Original induction”, 1 in “No Prime”). Furthermore, we eliminated trials according to the steps taken by Schul et al. (2004), see Appendix C, Table C.1 for further details).

Results

Data was analyzed using a 2 (Induction: Original Induction vs. No Induction) x 2 (Target: Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) mixed ANOVA with Induction as between subjects factor. No main effects and no interactions were found (all $F < 1$; see Figure 2, for means, *SD* and 95% CI see Appendix D, Table D.1). When analyzing only the two control conditions “No Face” and “No Prime”, none of the paired t-tests yielded significant effects, No Face: $t(23)=-.24$, $MSE=569.38$, $p=.810$, *g*[4] = -.013; No Prime: $t(22)=-.09$, $MSE= 2085.19$, $p = .928$, $g = .008$.[5]

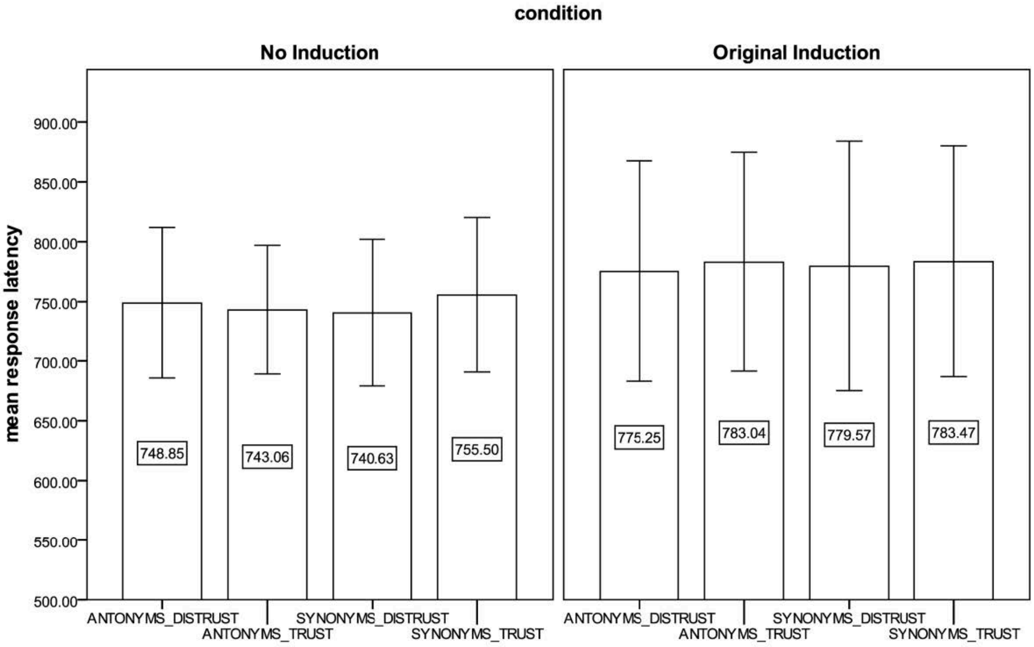


Figure 2:Mean response latency in ms and 95% confidence interval (CI) of “No Induction” and “Original Induction” in Experiment 1

Discussion

In Experiment 1, the hypotheses were not corroborated. Neither the interaction of the first experiment of Schul et al. (2004) in the condition “Original Induction” nor any effects in the condition “No Induction” were found. Due to intercultural differences in face perception (e.g., own race bias: Papesh et al., 2010), one explanation may be that the faces malfunctioned as a trust and distrust manipulation in the Central European cultural setting of Switzerland, compared to the cultural setting of Israel, where the original experiment was carried out. The pretest already gave a hint for this interpretation: although we found a good discrimination of untrustworthy and trustworthy faces, these differences were probably too small. Furthermore, we found lower untrustworthiness ratings of untrustworthy faces when shown with induction than without an induction. Finally, relatively high sadness ratings for trustworthy faces may be a reason for confounding the proposed effect. Therefore, we created new faces (Figure 3), using parameters for trustworthiness of Todorov et al. (2008).

Figure 3: Example of new created faces according to parameters of Todorov et al. (2008)

Another explanation for the failure may be the duration of the SOA that was originally used by Schul et al. (2004). The SOA of 82 ms and the ISI of zero ms are probably too short to guarantee a supraliminal perception of the prime. In general, priming effects using a short ISI (0-350 ms) are attributed to automatic spreading activation (Anderson, 1983), whereas effects with a longer ISI (over 400 ms) possibly appear due to expectation based strategies (Groot, Thomassen, & Hudson, 1986). Therefore, in Experiment 2, we increased the SOA up to 300 ms with an ISI of 218 ms, which is still in the range of automatic activation. The ISI consisted of a blank screen. The prime presentation time of 82 ms was held constant. Furthermore, we implemented a fixation cross at the start of each trial, to assure that attention was directed toward the stimuli. Finally, we added 80 trials to improve the reliability of the measures.

Experiment 2

Method

Design and participants. Experiment 2 was employed as a 2 (Target: Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) x 2 (Induction: Original Induction vs. No Induction) mixed design, using Target and Face as within-subjects factors. We added the control conditions “No Face” and “No Prime”. Participants were 99 students at the University of Bern, Switzerland, of which 60 participants were psychology undergraduates and 85 participants were women. The mean age of participants was 23.27

years (*SD* = 4.63). Participants were randomly assigned to each of the four between-subjects conditions (Original Induction: 24; No Induction: 25; No Face: 25; No Prime: 25, see Appendix A for power analyses) and they received either course credit for their participation or 5 CHF (5.16$ July 2013).

Materials and procedure. New coloured faces were created with Facegen Modeller version 3.5 (Singular Inversions, 2011; Figure 3). The faces were equally attractive, the ethnicity was Caucasian, and the symmetry of the face was held constant. The pictures had a size of 400 x 477 pixels with a grey background. In an online-pretest, 32 participants evaluated 21 faces of seven different face categories with varying parameters of trustworthiness (Todorov et al., 2008). We selected the two categories which discriminated well between a trustworthy and a neutral face, an untrustworthy and a neutral face, respectively (see Appendix B, Tables B.1 – B.3 for more details). Doubling the number of trials, we needed and thus created 80 untrustworthy and 80 trustworthy faces of these two categories.

In another pretest, like in Experiment 1, we tested the ratings of these selected faces and the influence of the induction on these ratings (see Appendix B, Figure B.2). The newly created faces discriminated also well between trustworthy and untrustworthy faces. Untrustworthy faces were rated more untrustworthy than trustworthy and trustworthy faces were rated more trustworthy than untrustworthy. The differences between these ratings of untrustworthiness and trustworthiness were larger than in Experiment 1. Additionally, the ratings for untrustworthy faces on untrustworthiness with induction was higher than without the induction, $t(40)=2.05$, $MSE= .91$, .$p=.047$, $d=.63$. This was also found for the trustworthiness of trustworthy faces, $t(40)=1.42$, $MSE=.68$, $p=.17$, $d = .44$. Furthermore, trustworthy faces were now judged less sad (Induction: $M=2.91$, $SD=1.07$; No Induction: $M=3.22$, $SD=0.97$) than the original trustworthy faces before. According to Todorov et al. (2008), trustworthy faces now received higher ratings on the parameter “happy” (Induction:

M=6.13, *SD*=0.69; No Induction: *M*=5.83, *SD*=0.93).

Sentences and adjectives were used for the induction, as in Experiment 1. For the priming phase with 160 trials, we also had to add 80 new word pairs (20 antonyms, 20 synonyms, 40 adjective-noun pairs). We added a fixation cross and changed the ISI from 0 ms to 218 ms.

Dependent variables.

All dependent variables were the same as in the previous experiment.

Eliminations. Data of 6 participants had to be excluded[6]. The other elimination steps were the same as in the previous experiment (see Appendix C, Table C.2 (160 trials) and C.3 (80 trials) for further details).

Results

Data was analyzed using a 2 (Target: Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) x 2 (Induction: Original Induction vs. No Induction) mixed ANOVA with Induction as between subjects factor. The three-way interaction effect was significant, $F(1,45)=12.225$, $MSE=2339.43$, $p=.001$, $\eta^2_p=.21$. When analyzing the condition “Original Induction” separately, no main effects were observed, both $F< 1.6$, but the hypothesized interaction was observed, $F(1, 21) = 4.44$, $MSE= 2270.99$, $p=.047$, $\eta^2_p =.17$. When paired with an untrustworthy face, antonyms were processed faster than when paired with a trustworthy face. Synonyms were processed faster with a trustworthy face, compared to a slower processing when paired with an untrustworthy face. In condition “No Induction”, no main effects were observed, both F<1.7, but an interaction effect appeared, $F(1, 24) = 8.19$, $MSE=2399.32$, $p = .009$, $\eta^2_p = .254$. The interaction, however, shows the opposite pattern than in the condition “Original Induction”: when paired with a trustworthy face, antonyms were processed faster than when paired with an untrustworthy face. Synonyms, in contrast, were processed faster with an untrustworthy face than with a trustworthy face (Figure 4). Analyzing only the two control conditions, none of the paired t-tests yielded significant effects, No Face: *t*(20) =-.44, *MSE*=708.54, *p* = .664, *g* = -.02; No Prime: *t*(24) = -1.49, *MSE*=557.48, *p* = .151, *g* = -.08; for means, *SD* and 95% CI see Appendix D, Table D.2 und D.3.

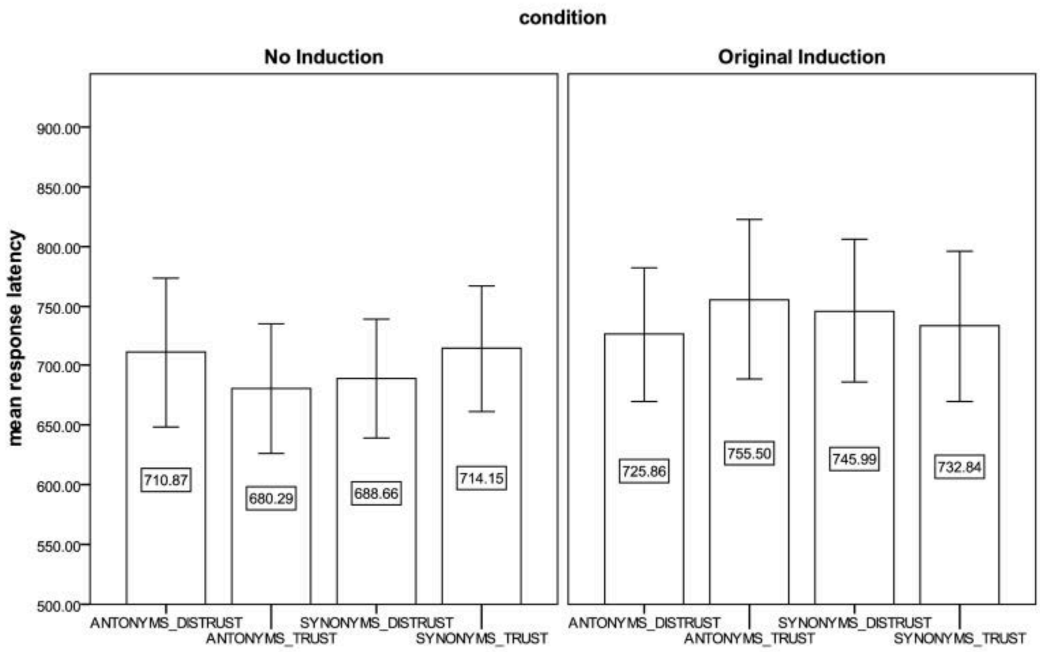


Figure 4: Mean response latency in ms and 95% CI of "No Induction" and "Original Induction" in Experiment 2

Post hoc analysis

In this analysis, we controlled post hoc for the canonicity of antonyms to receive first results regarding this aspect. Three interraters decided whether the antonyms were canonical or non-canonical[7]. In "Original Induction", using only canonical antonyms, a main effect for the "Target" variable, *F* (1, 21) = 11.82, *MSE* = 2129.62, *p* = .002, η^2_p = .36, and no interaction was found, all *F* < 2.2; *p* > .15. Canonical antonyms were generally processed faster than synonyms. With non-canonical antonyms, no main effect and no interaction was significant, *F* (1, 21) = 3.25, *MSE*=1247.75, *p* = .086, η^2_p = .13.

Using only canonical antonyms in "No Induction", we found a main effect for the "Target" variable, *F* (1, 24) = 7.80, *MSE*=3550.00, *p* = .010, η^2_p =.25, and the interaction was not significant, *F* (1, 24) = 3.09, *MSE*=2177.90, *p* = .092, η^2_p = .11. Numerically, especially under trust, antonyms were processed faster than synonyms. With non-canonical antonyms in "No Induction", no main effect and no interaction was significant, *F* (1, 24) = 3.22, *MSE*=2035.74, p = .086, η^2_p = .12. According to the means, antonyms were processed faster with a trustworthy face and synonyms were processed faster with an untrustworthy face (see Appendix D, Table D.4 for means and standard deviations).

Discussion

An introduction of European faces, an ISI of 218 ms, additional 80 trials and a fixation cross have led to results that are comparable to those of Schul et al. (2004). We found this comparable result in the condition “Original Induction”, where participants additionally learned that trustworthy faces go along with true statements and untrustworthy faces are accompanied with false statements. In the condition “No Induction”, however, the interaction had the opposite pattern: the processing of antonyms was faster with a trustworthy face and the processing of synonyms was faster with an untrustworthy face. This result contradicts the findings of Schul et al. (2004).

The question remains why in connection with the induction the interaction was comparable to the findings of Schul et al. (2004), whereas without the induction exactly the opposite result was demonstrated. In the following we will argue that the effects in the condition without induction show in an unbiased way that untrustworthiness may not generally enhance the detection of antonyms but can even hinder their detection, and that this problem may not appear in the condition with induction because the induction casts a damp over this problem.

First, it can be speculated that the induction itself is responsible for the comparable interaction effect. While Schul et al. (2004) consider that the induction phase only strengthens the genuine effect of trustworthy and untrustworthy faces, we think that it may be possible that trustworthy and untrustworthy faces themselves gain the function of primes during the induction, as Deutsch, Kordts-Freudinger, Gawronski & Strack (2009) formulated in their research. Participants learn to agree sentences presented with trustworthy faces and to deny sentences presented with untrustworthy faces. Trustworthy faces prime agreement, and untrustworthy faces prime denial. If a subject is primed in this manner, the processing of an antonym may be hampered with trustworthy faces because the priming of agreement blocks the association of the opposite word, i.e., the negation of the prime word. Likewise, the processing of a synonym may be hampered with untrustworthy faces because the priming of denial blocks the association of the same word, i.e., the affirmation of the prime word.

Second, the opposite interaction effect without induction strengthens the argument that perceiving an untrustworthy versus a trustworthy face may not necessarily lead to a faster processing of all kinds of antonyms. It is possible that especially non-canonical antonyms are more difficult to detect in the untrustworthy condition. If the prime-target pair is a non-canonical antonym, like “easy-difficult”, but a participant thinks another opposite like “complicated”, reaction time for the target “difficult” might increase and instead of being faster when perceiving an untrustworthy face, the participant gets slower. Such an incorrect association can extend the reaction time. To analyze whether the interaction with the opposite pattern in the condition “No Induction” is due to the ambiguity of antonyms and does not appear if non-canonical antonyms are used, one has to control for the canonicity of antonyms, as we already did in the post hoc analysis of the results of Experiment 2. This analysis gave us a first hint that the canonicity of antonyms may play a role in this paradigm.

We assume that in our third experiment a main effect for canonical antonyms will appear, showing their faster processing in general, and that this may overshadow any kind of possible interaction effect. For non-canonical antonyms, we are not able to make precise predictions due the explained different degrees of ambiguity of non-canonical antonyms.

Experiment 3

Method

Design and participants. Experiment 3 was employed as a 3 (Target: Canonical Antonyms vs. Non-canonical Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) x 2 (Induction: Original Induction vs. No induction) mixed design, using Target and Face as within-subjects factors. One additional control condition “No Induction &

No Face" was implemented. Participants were 80 students of the University of Erfurt, Germany, of which 51 participants were psychology undergraduates and 68 participants were female. The mean age was 22.84 (*SD* = 2.76). Participants were randomly assigned to one of three conditions (Original Induction: 27; No Induction: 27; No Induction & No Face: 26, see Appendix A, for power analyses). They received either course credit or 3 € (3.86$ July 2013) for their participation.

Materials and procedure. The same faces and trivia sentences as in Experiment 2 were used. 13 canonical antonyms were already in the adjective set of Experiment 2. To have the same number of canonical, non-canonical antonyms and synonyms, seven additional canonical antonyms were taken from Paradis et al. (2009). This resulted in 20 trials with canonical antonyms, 20 with non-canonical antonyms and 20 with synonyms, mixed with 60 trials where an adjective was followed by a noun. In total, the priming phase consisted of 120 trials[8].

Up to eight participants took part in each session; each participant assigned his or her own experimental booth. The monitors[9] were comparable to those in Bern. Procedure was the same as in the previous experiments. For the two conditions "No Induction" and "No Induction & No Face", the experiment only consisted of the priming phase.

Dependent variables. All dependent variables were the same as in Experiment 1 and 2.

Eliminations. Data of one participant had to be excluded in "No Induction & No Face"-Condition. The other eliminations steps are listed in Appendix C, Table C.4.

Results

Data was analyzed using a 3 (Target: Canonical Antonyms vs. Non-canonical Antonyms vs. Synonyms) x 2 (Face: Untrustworthy Face vs. Trustworthy Face) x 2 (Induction: Original Induction vs. No Induction) mixed ANOVA with Induction as between subjects factor. Only the main effect for "Target" was significant, F (2, 52) = 25.907, MSE= 2490.20, p<.001, η^2_p=.333, all other F<1.1. Separate analyses for each induction condition showed the same: “Original Induction”: (*F* (2, 52) = 8.98, *MSE*=2579.89, $p < .001$, η^2_p =.26; “No Induction”: *F* (2, 52) = 18.39, *MSE*= 2400.51, $p < .001$, η^2_p =.59. Even in the control condition “No Induction & No Face”, this main effect for “Target” was prevalent, F (2, 48) = 22.25, *MSE*=794.96, $p < .001$, η^2_p = .48; see Appendix D, Table D.5 for means, *SD* and 95% CI. Comparing only canonical antonyms with synonyms, the clear main effect is visible (Figure 5).

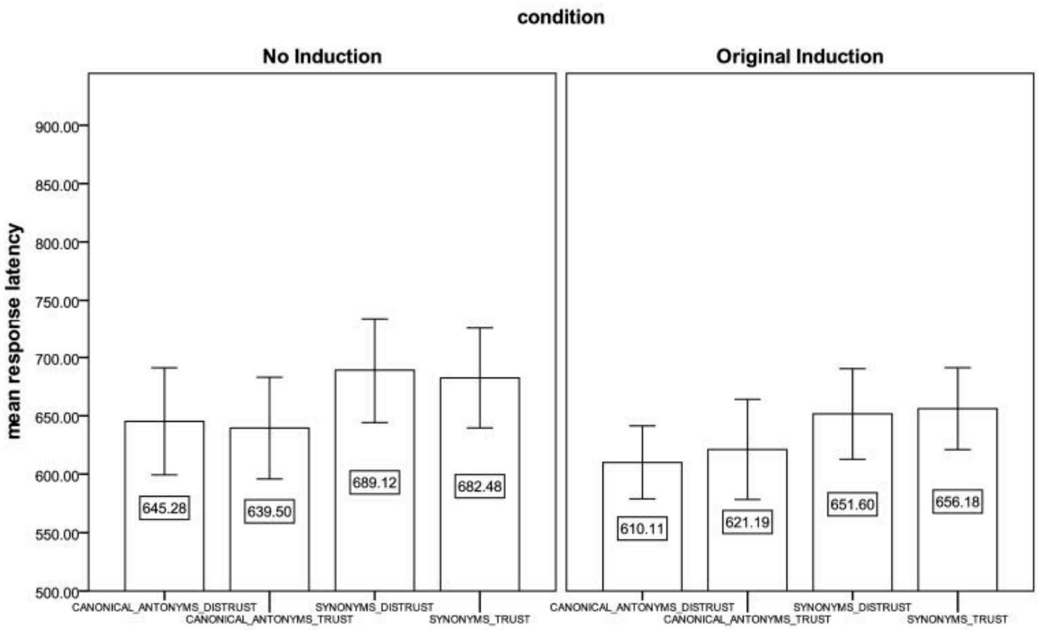


Figure 5: Mean response latency in ms of canonical antonyms versus synonyms in the conditions “No Induction” and “Original Induction” in Experiment 3

This main effect for “Target” disappeared when we analyzed the comparisons with non-canonical antonyms only (Figure 6). Neither a main effect nor an interaction appeared, independent of which condition, all $F < 1$, $t < 1$.

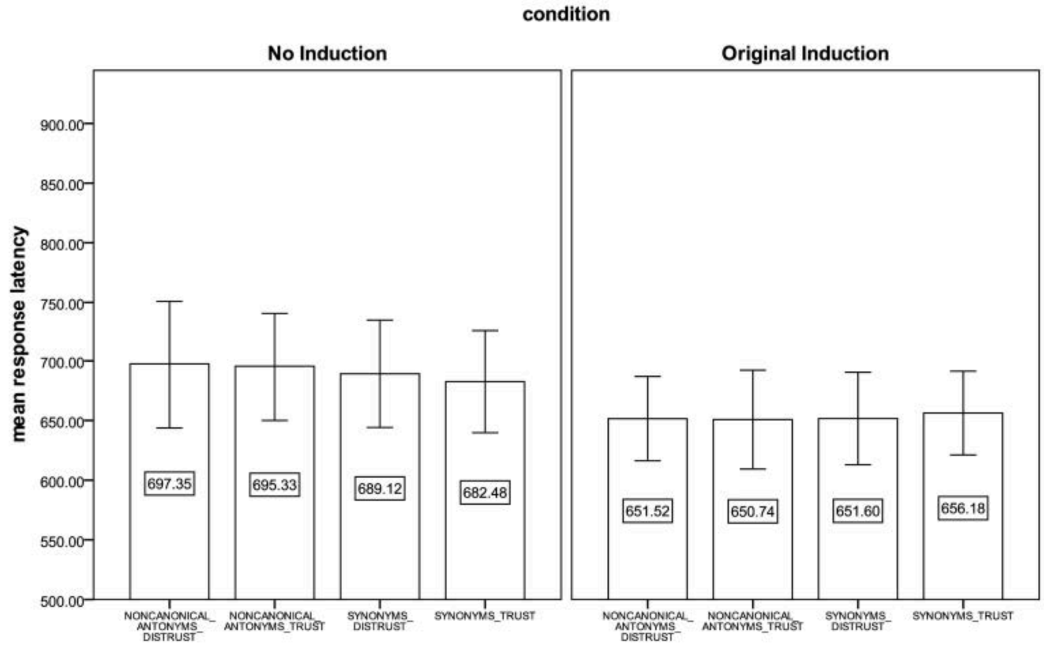


Figure 6: Mean response latency in ms of non-canonical antonyms versus synonyms in the conditions “No Induction” versus “Original Induction” in Experiment 3

Discussion

Designing the experimental task with the same number of canonical and non-canonical antonyms eliminates all effects of trustworthiness of faces. As hypothesized, canonical antonyms were processed faster than synonyms with and without the original induction phase, independent of presenting the target word with a trustworthy or an untrustworthy face. Analyzing the comparisons with non-canonical antonyms, one may speak of a “flat line”. No main effects and no interaction were found. Thus, the assumption that the interaction found by Schul et al. (2004) is only due to the application of the induction phase has not been corroborated. It seems that it has an influence, but that a thinking the opposite in this paradigm additionally depends on the set of used antonyms and their (non-) canonicity and the found interaction effects in the original experiment and our second experiment may only be the result by chance or by the proper composition of canonical and non-canonical antonyms.

General Discussion

The replication of the interaction as predicted by Schul et al. (2004) was successful in one of three experiments, and only if the induction phase took place. Without an induction, we found a contrary interaction effect in Experiment 2. But still we do not conclude that the theoretical assumption has been falsified that a state of distrust triggers a thinking the opposite. This conclusion would be drawn too early, because there are still some objections concerning the experimental paradigm.

Let us first elaborate on the assumed state of distrust and trust that should have been manipulated by the trustworthiness of faces. Trust- or untrustworthiness of a face or person, and observers‟ state of trust or distrust can not necessarily be equated (Winkielman, Knutson, Paulus, & Trujillo, 2007; Mayer et al., 1995). This is an important distinction, especially considering the fact that, in this experimental paradigm, participants perceived many trustworthy and untrustworthy faces one after another in a random order. Up to now, it remains an open question whether participants also changed their feelings of trust and distrust accordingly. It is possible that participants living in Israel can change their feelings of trust and distrust more rapidly than Swiss or Germans, because the former are much more sensitive towards insecurity (Bar-Tal & Jacobson, 1998) and, therefore, they distrust more easily. This serves as a possible explanation why we rarely replicated the results of Schul et al. (2004). Nevertheless, further experiments may prove that a rapid order of trustworthy and untrustworthy faces can actually lead to the respective states of trust and distrust. The social context and the behavior of the interaction partner (Tuk, Verlegh, Smidts, & Wigboldus, 2009) might make this possible.

Second, the conclusion of Schul et al. (2004) that distrust automatically enhances a thinking the opposite is questionable because of the experimental paradigm employed. At first sight, it seems appropriate to study thinking the opposite by testing whether untrustworthy faces lead to a faster detection of antonyms and trustworthy faces lead to a faster detection of synonyms. Conducting our experiments, however, we gradually came to the conclusion that antonyms are not ideal for testing a thinking the opposite because of the canonicity of antonyms.

As long as we use only canonical antonyms, it will be unambiguously clear which antonym should follow the prime. But because canonical antonyms are processed much faster than synonyms, a problem arises that this main effect may overshadow any possible interaction. As soon as non-canonical antonyms are included, however, one has to take into account that they differ substantially in their degree of ambiguity. If participants were to associate the appropriate antonym with a given adjective, we found that in the case of non-canonical antonyms up to twelve different answers (possible antonyms) were given[10]. Hence, the following problem becomes obvious: if participants have to detect canonical and non-canonical antonyms, the probability of having the correct association, varies substantially. Thus, if one uses non-canonical antonyms, it may be almost impossible to control whether participants have correct or incorrect associations with regard to the target word. This will be especially important in the condition of untrustworthy faces where participants are supposed to think the opposite faster, because any incorrect association may then extend their reaction time. The results of Experiment 2 showed that, without the induction, untrustworthy faces led to longer reaction times in detecting antonyms. Nevertheless, we have to analyze reasons why we could not find such an interaction in Experiment 3, where the canonicity of antonyms was controlled. In order to control the canonicity in this experiment, we had to enlarge the number of canonical antonyms and to modify some of the adjective pairs used previously. Thus, it is quite probable that the changing of the compilation of adjective pairs and particularly the specific combination of antonyms with various degrees of ambiguity, and not the trustworthiness of faces, has led to this result. However, this interpretation has to be tested in further experiments. But up to now, it becomes clear that canonicity plays a crucial role, if one is interested in analyzing thinking the opposite by means of the paradigm of antonyms and synonyms.

Our final objection concerns the role of the induction. While we were not able to replicate the interaction found by Schul et al. (2004) without the induction, we replicated it once in the condition with the original induction in Experiment 2. This result in Experiment 2 may have been achieved simply by chance, but in combination with the results of Weil (2010), there is at least some indication that the induction itself may be sufficient to produce the interaction effect. During the induction, which consists of 80 trials, clear associations between truth and a trustworthy face and between untruth and an untrustworthy face are created. Thus, faces, or even geometrical figures, as in Weil‟s experiment, become a reliable prime for knowing the truth or untruth instead for trust and distrust. Although distrust is normally qualified as a state of uncertainty (Lewicki, McAllister, & Bies, 1998), in the induction, participants learn certainty: they learn to be sure that they have to affirm statements of trustworthy faces and they learn to be sure that they have to negate statements of untrustworthy faces. The learned affirmation tendency of participants, when they perceived a trustworthy face, may have hampered their ability to recognize antonyms instead of facilitating the recognition of synonyms, as Schul et al. (2004) concluded. Likewise, the negation tendency of participants, when they perceived an untrustworthy face, may have obstructed their ability to recognize synonyms instead of facilitating the recognition of antonyms.

Limitations

One of the limitations of our experiments is the missing quadruplets in Experiment 3. There is not always a synonym, a non-canonical and canonical antonym from the same word, but from different ones. We have chosen this deliberately to avoid on one hand that participants have to decide more than once about the same word and on the other hand a canonical antonym does not have another antonym by definition. Furthermore, the role of canonicity has to be further analyzed in future studies. The process which antonym is associated has to be controlled more so that one may gain further insight what makes a thinking the opposite in this paradigm happen or not. Finally, some original adjective pairs were substituted because they were not translatable from Hebrew to proper contemporary German.

Conclusion

Despite these limitations, the three experiments conclusively show that a thinking the opposite under distrust cannot be adequately examined within the original paradigm. In order to positively confirm this, further research is needed to get more information about the real associations with regard to antonyms and about the role of the induction phase.

Word count main text including title and abstract: 6828

Acknowledgements

We thank Ruth Mayo for providing the original faces and adjectives and for giving us advice to analyze the data accordingly to their experiment. We thank Benedikt Bietenhard for translating the adjectives from Hebrew to German. Finally, we thank Mario Gollwitzer, Galina Savukova, Jörg Hupfeld-Heinemann and Frank Renkewitz for their helpful comments on previous drafts of this paper.

Appendix A

Power analysis

Original experiment

Result of a 2x2 ANOVA for repeated measures: F (1, 23) = 5.15, *p*<.05 (in the original article, p.673)

For such a 2x2 repeated measures design, hypothesizing a small effect of f„=.14 (corrected for the second within-factor, like Rasch, Friese, Hofmann & Naumann (2010) recommended) and a correlation of the measures of .9 and a minimal power of .80, 23 participants are required (Faul, Erdfelder, Lang, & Buchner, 2007)

Experiment 1 and 2

For a 2x2x2 ANOVA for repeated measures, a correlation between measures of .9, a small effect size of f‟=.14, corrected for second within factor and one between factor, and a power of 80%, 384 participants are required for the between effect, 23 for within effects and 24 for within-between interaction. Only analyzing the within-factors, 24 participants are required (Faul et al., 2007)

Experiment 3

For a 3x2x2 ANOVA for repeated measures, a correlation between measures of .9, a small effect size of f‟=.173, corrected for second within factor and one between factor, and a power of 80%, 252 participants are required for the between effect, 16 for within effects and 16 for within-between interaction (Faul et al., 2007)

Appendix B

Figure B.1 Pretest Experiment 1

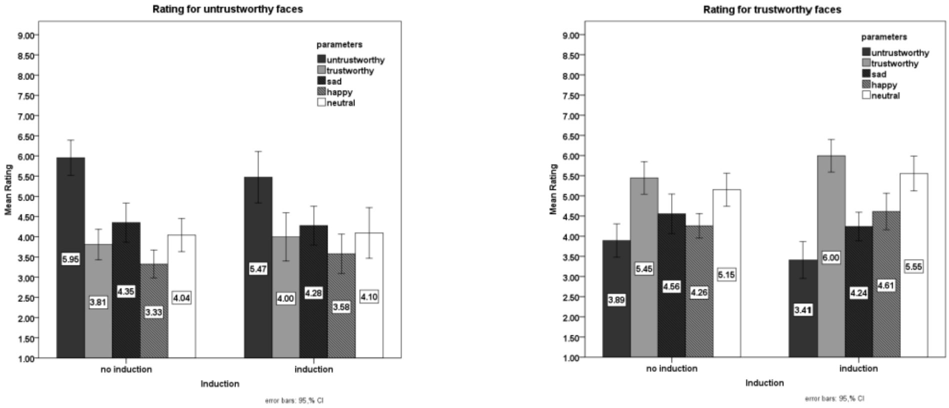


Table B.1 Pretest 1 Experiment 2

Parameters for faces in seven different categories.

	Parameters				
	Brow ridge inner	Cheek bones Shallow/ pronounced	Chin wide/thin	Nose sellion Shallow/deep	Eye squint left and right
untrustworthy 3	-6 SD	-6 SD	+6SD	+6SD	+1.00
untrustworthy 2	-4 SD	-4 SD	+4SD	+4SD	+1.00
untrustworthy 1	-2 SD	-2 SD	+2SD	+2SD	+1.00
neutral face	0 SD	0 SD	0 SD	0 SD	0.5
trustworthy 1	+2 SD	+2 SD	-2SD	-2SD	0.00
trustworthy 2	+4 S	+4 SD	-4SD	-4SD	0.00
trustworthy 3	+6 SD	+6 SD	-6SD	-6SD	0.00

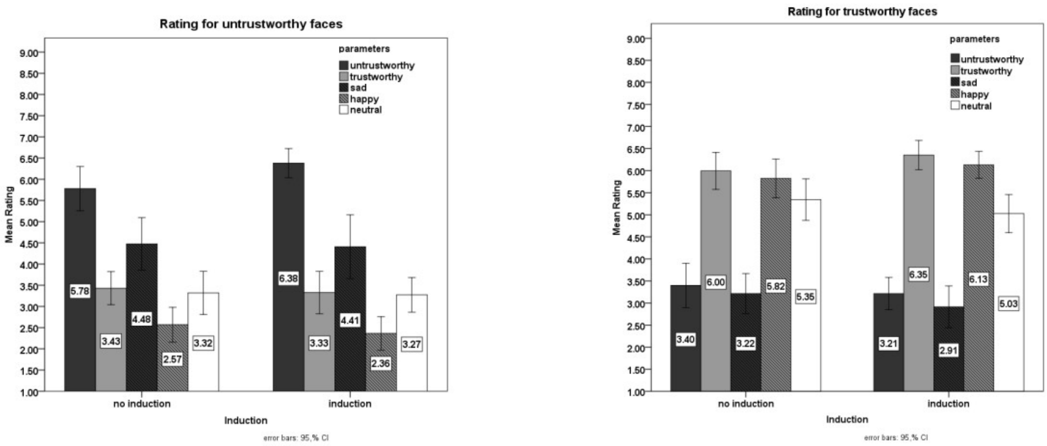
Table B.2 Pretest 1 Experiment 2 Number of faces in categories from untrustworthy to trustworthy.

category / kind of face	
untrustworthy 3	3
untrustworthy 2	3
untrustworthy 1	3
neutral face	3
trustworthy 1	3
trustworthy 2	3
trustworthy 3	3
Total presented faces	21

Table B.3 Pretest 1 Experiment 2

category / kind of face	Rating	
	Trust	Distrust
(*N*=32)	*M* (*SD*)	*M* (*SD*)
untrustworthy 3	28.75 (22.51)	69.77 (24.68)
untrustworthy 2	38.02 (18.08)	59.85 (21.77)
untrustworthy 1	42.49 (16.57)	53.79 (19.78)
neutral face	59.97 (14.72)	38.06 (16.07)
trustworthy 1	66.03 (15.89)	31.57 (17.52)
trustworthy 2	61.45 (14.69)	34.63 (14.61)
trustworthy 3	58.78 (16.63)	36.86 (18.55)

Figure B.2 Pretest 2 Experiment 2



Appendix C

Table C.1

Overview of Trial Eliminations in Experiment 1. Priming Accuracy and Number of Final Trials included.

	condition			
	original induction	no induction	no face	no prime
N	28	27	24	24
initial number of trials	2240	2160	1920	1920
participants with more than 20 % errors in the priming task	3 = 240 trials	0	0	1 = 80 trials
number of trials after eliminating participants with more than 20 % errors in the priming task = t1	2000	2160	1920	1840
failure to respond within the 2s response window	58	18	12	30
response latencies deviating more than 3 standard deviations from the mean of the individual subject	33	39	34	33
wrong answer	57	73	48	27
number of eliminated trials	148	130	94	90
final number of trials = t2	1852	2030	1826	1750
accuracy in the priming task	97.01%	96.53%	97.44%	98.48%
number of trials from t1	92.60%	93.98%	95.10%	95.11%

Table C.2

Overview of Trial Eliminations in Experiment 2 with 160 trials. Priming Accuracy and Number of Final Trials included.

	condition			
	original induction	no induction	no face	no prime
n	24	25	25	25
initial number of trials	3840	4000	4000	4000
participants with more than 20 % errors in the priming task	2 = 320 trials	0	4 = 640 trials	0
number of trials after eliminating participants with more than 20 % errors in the priming task = t1	3520	4000	3360	4000
failure to respond within the 2s response window	21	13	36	25
response latencies deviating more than 3 standard deviations from the mean of the individual subject	65	65	68	80
wrong answer	132	180	174	165
number of eliminated trials	218	258	278	270
final number of trials = t2	3302	3742	3082	3730
accuracy in the priming task	96.16%	95.41%	94.66%	95.76%
number of trials from t1	93.81%	93.55%	91.73%	93.25%

Table C.3
Overview of Trial Eliminations in Experiment 2 with 80 trials. Priming Accuracy and Number of Final Trials included.

	condition			
	original induction	no induction	no face	no prime
n	24	25	25	25
initial number of trials	1920	2000	2000	2000
participants with more than 20 % errors in the priming task	2 = 160 trials	1 = 80 trials	6 = 480 trials	0
number of trials after eliminating participants with more than 20 % errors in the priming task = t1	1760	1920	1520	2000
failure to respond within the 2s response window	18	11	10	21
response latencies deviating more than 3 standard deviations from the mean of the individual subject	32	30	34	40
wrong answer	71	93	92	95
number of eliminated trials	121	134	136	156
final number of trials = t2	1639	1786	1384	1844
accuracy in the priming task	95.85%	95.05%	93.77%	95.10%
number of trials from t1	93.13%	93.02%	91.05%	92.20%

Table C.4

Overview of Trial Eliminations in Experiment 3. Priming Accuracy and Number of Final Trials included.

	condition		
	original induction	no induction	no face and no induction
n	27	27	26
initial number of trials	3240	3240	3120
participants with more than 20 % errors in the priming task	0	0	1 = 120 trials
number of trials after eliminating participants with more than 20 % errors in the priming task = t1	3240	3240	3000
failure to respond within the 2s response window	19	22	9
response latencies deviating more than 3 standard deviations from the mean of the individual subject	78	59	59
wrong answer	173	175	152
number of eliminated trials	270	256	220
final number of trials = t2	2970	2984	2780
accuracy in the priming task	94.50%	94.46%	94.82%
number of trials from t1	91.67%	92.10%	92.67%

Appendix D

Descriptive Statistics for all three experiments (Tables D.1 – D.5).

Table D.1

Experiment 1: Means and (Standard Deviation) of Response Latency (in ms) and [95% Confidence Interval] in the Lexical Decision Task. .

condition	adjective target	adjective target	noun target
no prime (*n* =23)		no prime	no prime
untrustworthy faces		758.79 (128.47), [703.24, 814.34]	774.64 (133.14), [717.06,832.21]
trustworthy faces		760.03 (156.92), [692.17,827.89]	772.03 (137.39), [716.62,831.44]
no face (*n* =24)	synonym prime target pair	antonym prime target pair	irrelevant prime
polygons	752.00 (118.26), [702.06, 801.93]	750.32 (128.80), [695.93, 804.70]	780.85 (116.75), [731.55,755.18]
original induction (*n* =25)			
untrustworthy faces	779.57 (252.95), [675.16, 883.99]	775.25 (223.41), [683.03, 867.47]	810.52 (253.48), [731.55,755.18]
trustworthy faces	783.47 (233.90), [686.92, 880.01]	783.04 (222.01), [691.39, 874.67]	809.76 (243.20), [709.37,910.14]
no induction (*n* =27)			
untrustworthy faces	740.63 (155.41), [679.10, 802.11]	748.85 (159.70), [685.68, 812.03]	756.09 (124.62), [712.79,811.39]
trustworthy faces	755.50 (163.91). [690.66, 820.34]	743.06 (136.45), [689.08, 797.04]	764.41 (117.48), [717.94,810.88]

Table D.2

Experiment 2 (160 trials): Means and (Standard Deviation) of Response Latency (in ms) and [95% Confidence Interval] in the Lexical Decision Task.

condition	adjective target	adjective target	noun target
no prime (*n* =25)		no prime	no prime
untrustworthy faces		679.60(106.58), [637.82, 721.37]	701.87 (101.92), [661.92, 741.82]
trustworthy faces		689.51 (119.24), [642.77, 736.25]	687.48 (101.77), [647.59, 727.38]
no face (*n*=24)	synonym prime target pair	antonym prime target pair	irrelevant prime
polygons	679.82 (154.75), [613.64, 746.01]	676.20 (149.58), [612.23, 740.18]	688.82 (119.60), [637.67, 739.98]
original induction (*n* =22)			
untrustworthy faces	712.53 (125.40), [660.13, 764.93]	698.12 (126.94), [645.08, 751.16]	712.74 (110.75), [666.46, 759.02]
trustworthy faces	709.33 (127.33), [656.13, 762.54]	722.79 (136.76), [665.64, 779.94]	720.11 (109.47), [674.37, 765.85]
no induction (*n* =25)			
untrustworthy faces	664.43 (93.01), [627.97, 700.89]	674.78 (116.81), [628.99, 720.57]	686.57 (90.87), [650.95, 722.19]
trustworthy faces	684.14 (113.79), [639.53, 728.75]	656.01 (121.04), [608.57, 703.46]	687.15 (96.58), [649.29, 725.01]

Table D.3

Experiment 2 (80 trials): Means and (Standard Deviation) of Response Latency (in ms) and [95% Confidence Interval] in the Lexical Decision Task.

condition	adjective target	adjective target	noun target
no prime (*n* =25)		no prime	no prime
untrustworthy faces		709.13 (129.32), [658.44, 759.83]	730.40 (107.67), [688.19, 772.61]
trustworthy faces		700.18 (126.17), [650.72, 749.64]	717.88 (130.82), [666.60, 769.17]
no face (*n*=19)	synonym prime target pair	antonym prime target pair	irrelevant prime
polygons	710.89 (169.27), [634.78, 787.01]	702.64 (132.03), [643.27, 762.01]	717.42 (119.72), [663.59,771.25]
original induction (*n* =22)			
untrustworthy faces	744.18 (134.55), [687.96, 800.40]	737.85 (146.94), [676.45, 799.25]	752.07 (128.73), [698.28, 805.87]
trustworthy faces	735.55 (144.69), [675.09, 796.02]	766.07 (154.93), [701.33, 830.81]	758.72 (115.19), [710.58, 806.85]
no induction (*n* =25)			
untrustworthy faces	708.31 (123.42), [658.93, 757.69]	718.44 (158.35), [655.09,781.79]	713.02 (99.91), [673.04, 752.99]
trustworthy faces	726.32 (131.72), [673.62, 779.02]	690.18 (137.47), [635.18,745.18]	714.25 (132.39), [661.28, 767.22]

Table D.4

Experiment 2 post-hoc analysis: Mean and (Standard Deviation) of Response Latency (in ms) and [95 % Confidence Interval] in the Lexical Decision Task.

condition	synonym prime target pair	canonical antonym prime target pair	non-canonical antonym prime target pair
original induction (*n* =22)			
untrustworthy faces	712.53 (125.40); [660.13, 764.93]	662.22 (128.42); [608.56, 715.88]	715.79 (133.05); [660.20, 771.39]
trustworthy faces	709.33 (127.33); [656.13, 762.54]	692.00 (1156.13); [626.76, 757.24]	739.76 (141.34); [680.70, 798.82]
no induction (*n* =25)			
untrustworthy faces	664.43 (93.01); [627.97, 700.89]	647.54 (133.46); [595.22, 699.86]	682.32 (115.07); [637.21, 727.43]
trustworthy faces	684.14 (113.79); [639.53; 728.75]	634.45 (125.41); [585.29; 683.61]	669.67 (121.57); [622.01, 717.32]

Table D.5

Experiment 3: Means and (Standard Deviation) of Response Latency (in ms) and [95% Confidence Interval] in the Lexical Decision Task.

condition	adjective target	adjective target	adjective target	noun target
no face / no induction (*n* =25)	synonym prime target pair	canonical antonym prime target pair	non-canonical antonym prime target pair	irrelevant prime
polygons	657.75 (87.10), [621.79, 693.70]	610.98 (87.45), [574.88, 647.08]	656.31 (85.37), [621.08, 691.55]	652.56 (78.69), [620.08,685.04]
original induction (*n* =27)				
untrustworthy faces	651.60 (97.97), [612.84, 690.35]	610.11 (78.94) [578.88, 641.33]	651.52 (89.37), [616.16, 686.87]	648.00 (72.78), [620.21,677.79]
trustworthy faces	656.18 (88.57), [621.14, 691.21]	621.19 (108.39), [578.31, 734.41]	650.74 (104.68), [609.33, 692.15]	643.29 (76.30), [613.11,673.47]
no induction (*n* =27)				
untrustworthy faces	689.12 (113.57), [644.19, 734.04]	645.28 (115.85), [599.45, 691.11]	697.35 (135.47), [643.76, 750.94]	674.63 (75.95), [644.59,704.67]
trustworthy faces	682.48 (108.21), [639.67, 725.28]	639.50 (110.11), [595.95, 664.07]	695.33 (114.77), [649.93, 740.73]	668.18 (75.36), [638.37,697.99]

References

Anderson, J. R. (1983). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.

Bar-Tal, D. & Jacobson, D. (1998). The Elusive Concept of Security and Its Expression in Israel. In D. Bar-Tal, D. Jacobson, A. Klieman (Eds.), *Concerned with Security: Learning from the Experience of Israeli Society.* Greenwich, CT: JAI.

Borenstein, M., Hedges, L.V., Higgins, J.P. & Rothstein, H.R. (Eds.). (2009*). Introduction to Meta-Analysis.* West Sussex, UK: John Wiley & Sons, Ltd.

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and Fragile: A New Look at the Automaticity of Negation Processing. *Experimental Psychology*, *56*(6), 434–446. doi:10.1027/1618-3169.56.6.434

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17695343

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 publicity and inadmissible testimony. *Personality and Social Psychology Bulletin*, *23*, 1215–1226. doi:10.1177/01461672972311008

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*(6), 1164–1184. doi:10.1037/0022-3514.70.6.1164

Friesen, J., & Sinclair, L. (2011). Distrust and Simultaneous Activation of Multiple Categories. *Social Psychological and Personality Science*, *2*(1), 112–118. doi:10.1177/1948550610382666

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*(2), 127–171. doi:10.1016/0010-0277(92)90060-U

Gilbert, D. T. (1993). Attribution and Interpersonal Perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99–147). New York: McGraw Hill.

Gorn, G. J., Jiang, Y., & Johar, G. V. (2008). Babyfaces, Trait Inferences, and Company Evaluations in a Public Relations Crisis. *Journal of Consumer Research*, *35*(1), 36–49. doi:10.1086/529533

Groot, A. M. de, Thomassen, A. J., & Hudson, P. T. (1986). Primed-lexical decision: The effect of varying the stimulus-onset asynchrony of prime and target. *Acta Psychologica*, *61*(1), 17–36. doi:10.1016/0001-6918(86)90019-3

Lewicki, R. J., & Bunker, B. B. (1996). Developing and Maintaining Trust in Work Relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations. Frontiers of Theory and Research* (pp. 114–139). Thousand Oaks, California: Sage Publications.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and Distrust: New Relationships and Realities. *Academy of Management Review*, *23*(3), 438–458. Retrieved from http://www.jstor.org/stable/259288

Mayer, J., & Mussweiler, T. (2011). Suspicious spirits, flexible minds: When distrust enhances creativity. *Journal of Personality and Social Psychology*, *101*(6), 1262–1277. doi:10.1037/a0024407

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734. Retrieved from http://www.jstor.org/stable/258792

Millisecond. (2012). Inquisit 3.0.6.0. Seattle, WA. Retrieved from http://www.millisecond.com/

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition*, *116*(2), 283–288. doi:10.1016/j.cognition.2010.05.001

Paradis, C., Willners, C., & Jones, S. (2009). Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon*, *4*, 380–429. doi:10.1075/ml.4.3.04par

Rasch, Friese, Hofmann & Naumann (2010). *Quantitative Methoden. Band 2* (3. Auflage). Heidelberg: Springer.

Schul, Y. (1993). When warning succeeds: The effect of warning on success of ignoring invalid information. *Journal of Experimental Social Psychology*, *29*, 42–62. doi:10.1006/jesp.1993.1003

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. doi:10.1006/jesp.1996.0011

Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding Under Trust and Distrust: The Spontaneous Activation of Incongruent Cognitions. *Journal of Personality and Social Psychology*, *86*(5), 668–679. doi:10.1037/0022-3514.86.5.668

Schul, Y., Mayo, R., & Burnstein, E. (2008). The value of distrust. *Journal of Experimental Social Psychology*, *44*(5), 1293–1302. doi:10.1016/j.jesp.2008.05.003

Sedlmeier, P., & Renkewitz, F. (2008). *Forschungsmethoden und Statistik in der Psychologie.* München [u.a.]: Pearson Studium.

Singular Inversions. (2011). FaceGen Modeller 3.5. Toronto, CAN. Retrieved from http://facegen.com

Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth Biases and Aroused Suspicion in Relational Deception. *Communication Research*, *19*(3), 326–345. doi:10.1177/009365092019003002

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience, 3*(2), 119–127. doi:10.1093/scan/nsn009

Tuk, M. A., Verlegh, P. W. J., Smidts, A. L. E., & Wigboldus, D. H. J. (2009). Interpersonal relationships moderate the effect of faces on person judgments. *European Journal of Social Psychology, 39,* 757–767. doi:10.1002/ejsp

Weil, R. (2010). *Die Antonymie-Heuristik: Automatische Falsifikation Valenter Information* (Dissertation). Universität, Trier.

Wentura, D., & Degner, J. (2010). A Practical Guide of Sequential Priming and Related Tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York, NY: Guilford Press.

Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J. L. (2007). Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology*, *11*(2), 179–192. doi:10.1037/1089-2680.11.2.179

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

Word count references: 864

Footnotes

[1] We thank Ruth Mayo for her comments on this aspect, as we presented her our experiment.
[2] Flat Panel Monitor, Panel Size: 17", Aspect Ratio 5:4, Max Resolution 1280 x 1024 at 60 Hz
[3] The induction phase was repeated once.

[4] Formula for effect sizes out of t-test results for dependent measures: s $\frac{\sqrt{}}{}$ (Sedlmeier & Renkewitz (2008), p.421

[6] We report the eliminations and results for 160 trials.
[7] We decided to identify an adjective pair as canonical, if at least two of three interraters rated them as canonical.
[8] We had to reduce the trials from 160 to 120 because it was difficult to find more than 20 canonical antonyms.
[9] Flat Panel Monitor, Panel Size: 19", Aspect Ratio 4:3, Max Resolution 1280 x 1024 at 60 Hz
[10] Pre-tests which were conducted for other experiments at our division showed that an adjective might have up to 12 different antonyms.