Distrust Facilitates Analytical Reasoning

Nicole Ruffieux [1*] and Margit E. Oswald[1]

[1]Institute for Psychology, Social Psychology, University of Bern, Switzerland; E-mail addresses: 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 address: nicole.ruffieux@psy.unibe.ch (N. Ruffieux); Phone: 0041/31/6314724, Fax: 0041/31/6318212

Abstract

We examined whether distrust (vs. neutral condition) triggers non-routine, analytical reasoning processes and improves the accuracy of deception detection. We conducted two experiments to investigate the influence of distrust on analytical thinking. In the first experiment, participants had to determine whether the reports are true or falsified. The results show that distrust improves the accuracy of detecting lies without reducing truth accuracy. Qualitative analyses of participants' reported reasons of their veracity judgments suggest that participants under distrust rely especially on the logical consistency of arguments. In the second experiment, we tested the assumption that distrust enhances analytical reasoning more straightforward. We applied the paradigm of belief bias and expected that participants under distrust (vs. neutral condition) would rely more on logic than on plausibility in their judgment of the validity of a conclusion. The results of the second experiment strongly support the assumption that distrust triggers and fosters analytical reasoning.

(150 words)

Keywords: distrust, deception detection, belief bias, analytical reasoning, subliminal priming, deductive reasoning

Distrust Facilitates Analytical Thinking

While trust is widely recognized as having mainly positive consequences, such as personal well-being or improvement of interpersonal relations (Mayer, Davis, & Schoorman, 1995; Kramer, Brewer, & Hanna, 1996; Diener & Seligman, 2004; Sitkin & George, 2005), distrust has often been associated with negative consequences, such as hostility towards others (Chambers & Melnyk, 2006; Sitkin & Stickel, 1996), lack of cooperation (Axelrod, 1984; Chou, 2006; Deutsch, 1960), or avoidance of interaction (Bies & Tripp, 1996; March & Olsen, 1975). However, especially a situation-specific distrust or suspicion could have also advantages as we will see.

If people are in a state of distrust, they have negative expectations regarding the intentions and behaviour of another person or even a group of people (Lewicki, McAllister, & Bies, 1998). While trust means renouncing social control (Mayer et al., 1995), a state of distrust makes it essential to control behaviour and intentions of the distrusted person. Promises or arguments in general are not taken for granted like in a state of trust. One of the most relevant functions of distrust is that people are not guided any longer by their first impression or the seemingly obvious reasons for another's behaviour, but rather take a closer look at the current circumstances. Thus, distrust can be interpreted as the tendency of people to be ready to resist the persuasive intent of others and to think about possible alternative explanations, so-called counter scenarios, of the obviously given interpretations (Fein, 1996; Schul, Burnstein, & Bardi, 1996; Schul, Mayo, & Burnstein, 2004; Schul, Mayo, & Burnstein, 2008). In accordance with these considerations, recent studies that induced distrust explicitly or by means of a subliminal priming paradigm associate distrust with social categorization in the way that it promotes multiple categorizations of people (Friesen & Sinclair, 2011). In addition, Mayer and Mussweiler (2011) demonstrated that induced distrust leads to an increase in creative thinking and cognitive flexibility. Furthermore, Posten and Mussweiler (2013) show evidence for a stereotype-reducing effect based on enhanced non-routine thought

processes under distrust. These advantages of a state of distrust are interesting and worth pursuing further.

Particularly from an evolutionary point of view, one should expect that a state of distrust has also the positive function of enhancing deception detection and the recognition of cheating (Gigerenzer & Hug, 1992). Survival in former times, such as the securing of food storage, depended significantly more than nowadays on the social cooperation among members of one's own group and the possibility of relying on oral agreements (Brewer, 2001; Gigerenzer et al., 1992). However, people often run the risk that someone is lying to them, distorting a fact when reproducing it, or perhaps entirely omitting an important detail. The relatively high prevalence of deception and "lie-telling" is well documented in research (e.g. DePaulo & Kashy, 1998; DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). Our ability to discern truth from falsity does not seem to be very far-reaching, however. Laypeople reach an overall accuracy rate of 54%, whereas simply guessing would yield an accuracy rate of 50% (Bond & DePaulo, 2006; Hartwig & Bond, 2014). Due to the given risk of being deceived and lied to, in combination with our marginal ability to know whether we can bestow faith to someone, a good portion of distrust can be seen as healthy and, in some extreme cases, even important for survival.

Although it would be important to know whether distrust indeed has the function to improve the veracity of deception detection, this question has yet been only scarcely examined. Miller, Mongeau, & Sleight (1986) suggest that suspicion positively affects accuracy in detecting deception. Recently, Reinhard and Schwarz (2012) have shown that the accuracy of lie detection can be improved if people elaborate the content of a message instead of relying on non-verbal behaviour. It improves the accuracy of lie detection because content cues, e.g., number of reported details, or logical consistency, are generally more valid than non-verbal cues, like gaze aversion (Vrij, 2008). However, the elaboration of information

processing in the study of Reinhard et al. (2012) was due to a manipulated negative mood rather than to a state of distrust.

In our study, we want to examine whether distrust triggers non-routine thought processes and improves deception detection. We assume that the reasoning process of distrusting people will be less guided by plausibility considerations, as it is usually is the case, and will instead take a closer look at the logical consistency of statements. Thus, the increased inclination to examine logical inconsistencies should improve veracity judgments because "inconsistency" is a rather valid cue to distinguish truth from lies (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, 2003). To say it in a nutshell: We assume that a state of distrust triggers analytical reasoning and will enhance the overall accuracy of discerning truth from lies.

The present study

In the first experiment, we examine whether a state of distrust (versus a control condition) enhances deception detection and whether participants under distrust are more convinced that their veracity judgments were driven by the logical structure of the arguments rather than the plausibility of the given arguments. In a second experiment, we will investigate our hypothesis that distrust enhances analytical reasoning by using the belief bias as a paradigm. The increase in non-routine thinking, thus not being misled by plausibility considerations, should decrease the belief bias. The belief bias that emerges from being misled by the plausibility of a conclusion should decrease under distrust, while analytical reasoning should increase.

Experiment 1

Empirical Predictions

We expect that participants in a mind-set of distrust will detect falsified statements significantly better than participants in a neutral mind-set. However, we do not assume that distrust simply increases a so-called lie bias, which is an improved accuracy of detecting lies

at the expense of correctly detecting the truth. We assume instead that people under distrust should have a better overall accuracy rate of deception detecting than people in a neutral condition because they are not only better at detecting lies, but are at the same time at least equally good at detecting the truth as participants in a neutral mind-set. We predict furthermore, that a qualitative analysis will show that participants in the distrust (versus control) condition report more often that their veracity judgments were based on considerations of logical consistency.

Method

Subjects and design. The study involved 60 students (40 female) aged 19 to 44 years (M = 23.67, SD = 4.29) from the University of Bern. We excluded three participants in advance from the subsequent analysis. One person did not solve the task at all, and the other two showed a very high error rate (>.83).

The study consisted of a 2 x 2 mixed subject design with the between-factor, "mind-set", (distrust/neutral) and the within-factor, the "reality status" of two statements (true/falsified). The subjects were randomly assigned to one of the two conditions.

Procedure and materials. The investigation was entirely computer-based. At the beginning, the researchers welcomed the participants and directed them to their seats. They immediately set off individually with the experiment on their respective computers. We tested 1-4 subjects per session. The subjects signed informed consent forms and were advised that they could stop the experiment at any time. Subsequently, the investigators started the experiment.

***Distrust Induction*.** We decided to manipulate distrust subliminally because of two reasons. First, an explicit manipulation of distrust may provoke counter-reactions of participants in order to control biasing feelings of distrust (Wegener & Petty, 1997; Strahan, Spencer, & Zanna, 2002). This may be especially true if participants are asked to judge the veracity of statements, like in our experiment. Second, former studies could successfully show that a

subliminal priming of distrust has the expected impact on individual information processing (Mayer et al., 2011; Posten & Mussweiler, 2011), and subsequently alters intentions and behaviour of a person (e.g. Mayer et al., 2011; Friesen et al., 2011; Légal, Chappé, Coiffard, & Villard-Forest, 2012). To induce distrust, we used a masked subliminal priming procedure. Participants proceeded a lexical decision task (LDT), after they had seen the welcome screen, which briefly explained what tasks they could expect and how long the investigation would take. In the LDT participants had to determine whether a letter string is a word or not a word. The subliminal priming consisted of two blocks of 24 trials, plus 2 practice trials at the beginning of the blocks. As subliminal prime, the word "distrust" (*Misstrauen*) appeared. The participants in the neutral condition also solved a word task, but without a hidden subliminal priming (blank priming). The subliminal prime was presented for 13 ms and was embedded in a forward mask of 3000 ms and a backward mask of 506 ms. The mask consisted of a string of letters ("XWXXWWXXWX") and entirely masked the prime. The temporal sequence of prime and target complies with the recommendations of Bargh and Chartrand (2000). We successfully pretested this manipulation. We used implicit and explicit measures for the manipulation check. The results of the implicit measure prove that the subliminal priming of distrust raises implicit distrust but without raising negative affect. Moreover, participants in the distrust condition seem to have a stronger feeling of distrust (explicit measure) than participants in the neutral condition (for details see Appendix A).

Instruction. After the LDT, participants read the following instructions:

> "Hereafter you will read two reports which were made by dog owners about an incident with their dogs. Afterwards we would like to know your opinion about this incident. Your task is therefore to address yourself critically, but neutrally to the two reports and the descriptions contained therein, and then indicate whether they are true."

Reading and Evaluation of Reports on Incidents. Following the instructions, the participants read a true and a falsified report. Steinbach (2011) provided the reports. The reports came from dog owners and contained descriptions of an incident in which their dogs

bit another person or another dog. The dog owners were asked to tell two different versions of an incident. Once, they were supposed to tell the truth without leaving out any important details. Another time, they were asked to describe their experiences as though a police officer was interviewing them. In this case, the incidents should be falsified in order to prevent possible legal consequences, e.g. being obliged to keep the dog on a lead. The resulting reports were transcribed and coded on the basis of the criteria-based content analysis (CBCA) (Steller & Köhnken, 1989). For Experiment 1, a true and a falsified report from different dog owners were selected that were almost equivalent in length (true report = 322 words, falsified report = 310 words), and in the severity of the reported incident. Both reports were defined as high stake incidents, which means that each report contained a dog bite with an injury. The reality status (true or falsified) of the reports served as a within factor (true/falsified), and their order was randomized across subjects.

Credibility judgment and judgment of truth. After each of the two randomly presented reports, participants had to indicate how credible they deemed it on a scale from 1 (not at all credible) to 5 (very credible). Additionally, they were asked to answer the forced-choice question whether the report was either true or falsified.

Reported cues of deception detection. After participants have made their veracity judgments we asked them an open question to find out *why they came to the judgment* that no, one or both reports were true respectively falsified. Participants could describe as many cues as necessary, e.g. "very detailed" or "contradictory", that assumingly guided their judgments and had to add after each cue, whether the actual value of the cue hints for a true or a falsified report.

Control questions. Finally, subjects answered control questions. We were interested in how participants felt while reading the reports (on a scale from -50 (bad) to + 50 (good)), how easy it was for them to come up with a judgment (scale of 1 (very difficult) – 6 (very easy))

and how sure they were about their judgments (scale of 1 (very insecure) – 6 (very secure)). Additionally, we asked them again how they felt at that moment.

Results

The following dependent variables were of interest to us: The accuracy of judgments of deception detection, the mean of the credibility judgments, and the frequency of the cues of deception detection reported by participants in the open question. We had preliminarily coded the specified cues, however (see below).

Accuracy of judgments. The overall accuracy rate under a state of distrust amounts to 76.67%, in comparison to 63.33% for participants in a neutral state: $F(1,58) = 1.2.90$, $p = .094$, $\eta_p^2 = .048$. Thus, the expected main effect of distrust was not significant. To gain more knowledge about whether distrust probably only increases the accuracy of lie detection *at the expense* of the correctly detected truth, we conducted separate analyses.

For the falsified report, we found a significant difference caused by mind-set in the percentage of correctly identifying a lie: χ^2 (1, $N = 60$) = 5.41, p =.020, V= .30, (1 – β err prob) = .64. Under a state of distrust, 66.70% were able to detect the falsified report correctly, whereas only 36.71% of participants in a neutral state could do so (Figure 1).

For the true report, however, the percentage of participants judging it correctly as true was as high in the distrust condition as it was in the control condition: χ^2 (1, $N = 60$) = .218, p = .64.V =.06, (1 – β err prob) = .052. For the true report, there was an unexpected high percentage of correct judgments for both mind-sets: 90.32% for the distrustful group and 90.00% for the neutral group.

<< Figure 1 >>

Credibility judgments. An ANOVA with repeated measures on the variable reality status of the report as the within factor and mind-set as the between factor revealed a

significant main effect for the reality status of the report: $F(1,58) = 40.83$, $p < .001$, $\eta_p^2 = .48$. Participants judged the true report ($M = 4.03$; $SD = .1.03$) as significantly more credible than the falsified report ($M = 2.87$; $SD = .76$). The main effect of mind-set, $F(1, 58) = 1.89$, $p = .17$, $\eta_p^2 = .032$, and the interaction of the reality status of the report x mind-set, $F(1,58) = 1.09$, $p = .30$, $\eta_p^2 = .018$, were not significant.

Another ANOVA with repeated measures showed that there are no significant differences in the latencies for making a judgment: all $F < 1$, $\eta_p^2 < .016$.

Coding of the specified cues of deception detection. In the next step, we conducted a content analysis of all cues of deception detection that were specified by participants in the open-ended question. Although people are not very good at knowing which cues of the reports have indeed triggered their veracity judgments (Park, Levine, McCornack, Morrison, & Ferrara, 2002; Hartwig et al., 2014), their answers may nevertheless be a hint about which cues might have guided their reasoning process during their judgments. The coding proceeded in several steps (Früh, 2011). First, we excluded those answers, like e.g. general comments on the reported incidents or on participants' own experiences with dogs, which were irrelevant to the research question. Additionally, we deleted all filler words. Second, according to the flow model of Mayring (2008), we defined each category as precisely as possible, so that a subsequent assignment of the participant's statements was possible. To facilitate this process, so-called "anchor examples" were selected from the statements. These examples are supposed to illustrate concisely, which statements should be assigned to which category. The four categories that we defined and used for the classification of the answers were Contradiction/Consistency, Plausibility/Believability, Narrative Style and Other Cues. Third, we summed up similar information into the four categories and conducted a frequency analysis. The categorization of cues happened to be independent of whether a particular cue was assumed to indicate true or falsified reports. Thus, a cue like "many details" was

subsumed under the category Narrative Style, as was the cue "too little details". Table 1 gives an overview of the frequency of the categorized responses in the two mind-sets.

<< Table 1 >>

The frequencies by which the categorized cues were used to identify a report as falsified or true differed significantly between the two mind-set conditions: χ^2 (3, N = 60) = 13.58, p = .004, V= 0.27. Participants under the condition of distrust answered that they had predominantly used the contradiction/consistency of the report as a cue (53.30%). In the control condition, participants primarily used plausibility/believability as the dominant cue (40.00%) (see Figure 2).

<< Figure 2>>

Control questions. Mood of the participants while reading the report did not differ between the two mind-sets: $F(1,58) = .76$, $p = .39$, $\eta_p^2 = .013$. Mood while reading the report showed no significant correlation with the overall accuracy rate in the dichotomous judgment: $r(60) = -.06$, $p = .67$. Likewise, perceived ease of the judgment and security of the judgment did not differ between mind-sets: $F(1,58) = .00$, $p = 1.00$, $\eta_p^2 = .00$; $F(1,58) = .308$, $p = .58$, $\eta_p^2 = .005$. These additional control variables showed again, no significant correlation with the overall accuracy in the dichotomous judgment: r (60) = -. 011, p =.93; r (60) = -. 057, p =. 67.

Discussion

It was one of our main interest to investigate whether distrust improves the accuracy of deception detection such that the accuracy of lie detection increases without a decrease of correctly detecting the truth. But we were also interested to receive first hints about whether

under distrust the veracity judgments is guided by different, more valid cues than under a neutral condition.

Accuracy and credibility. The overall accuracy of deception detection in the distrust condition was higher than in the control condition although the difference was not significant. A closer look at the separate accuracy rates of the true and the falsified report shows that there was a ceiling effect of the true report. The high percentage of participants correctly identifying the true report in both mind-sets (about 90%) made it difficult to reach a significant difference of the overall accuracy between the experimental and the control group. The true report obviously contained many cues that facilitated the veracity judgment. Thus, according to the independent coding of the deception cues (CBCA), the true report was not only plausible, but also logically consistent. Probably more important than the main effect of distrust on accuracy of deception detection is therefore the result of separate analyses: Distrusting participants were not only as good as participants in the control condition were at detecting the truth, but even significantly better at detecting the lie. Thus, although the overall accuracy was not significantly better, we could verify that distrust improved the detection of a lie, but not at the expense of detecting the truth. As expected, participants in a state of distrust do not generally evaluate the reports as less credible than participants in a neutral state.

Reported cues of deception detection. Depending on the mind-set, participants reported different cues that may have guided their veracity judgments. If they were in a state of distrust, they seemed to pay significantly more attention to logical contradictions/inconsistencies in order to reach a judgment. Participants in a neutral state, however, mentioned this cue as well, but seemed to rely primarily upon the plausibility/believability of the report. Thus, while plausibility/believability serves as a stronger cue in the control condition, participants in the distrust condition seemed to rely almost exclusively on the logical consistency of a report.

If a report is logically consistent, but not plausible, or the other way around, participants in a neutral state may have difficulties coming to a correct judgment. Participants in a state of distrust, however, who assumingly rely more on the logical consistency of a report, are misled less frequently by its pure plausibility. In Experiment 2, we intend to test this assumption.

Limitations

As self-reports are not necessarily a valid method of proving cognitive processes, the content analysis of participants' self-reports may only give a hint that distrust potentially facilitates analytical reasoning, that is, the use of contradictions or inconsistencies between statements as a cue in the veracity judgment. To examine directly and with more rigour, whether participants normally prefer plausibility as a cue, while they show an improved consideration of logical consistency under distrust, it would be necessary to vary logical consistency and plausibility independently of the reality status of different reports. However, logical consistency is one of the most valid cues to detect the truth of statements. Thus, logical consistency is confounded with the reality status of ecologically valid reports, and it will be quite difficult to find comparable reports where reality status and consistency vary independently. In Experiment 2, we therefore changed the paradigm to bypass the limitations mentioned above.

Experiment 2

Participants in the control condition of Experiment 1 seem to place more emphasis on belief/experience and less on logic while judging the veracity of the statements. They probably show a heuristic that is known as the "belief bias" (Wilkins, 1928; Morgan & Morton, 1944; Evans, Barston, & Pollard, 1983; Klauer, Musch, & Naumer, 2000). The belief bias is the tendency to judge the validity of conclusions based on the plausibility/believability of the arguments than on their logical structures (Evans et al., 1983).

To examine whether participants under distrust rely more on logical considerations and thus have a less-pronounced belief bias than people normally have, we conducted another experiment. Unlike in Experiment 1, where the veracity of reported incidents had to be judged, in Experiment 2, participants have to solve syllogisms, a task that researchers have often used to examine the belief bias. A syllogism consists of two interrelated assumptions, called a major premise and a minor premise, that lead to a conclusion. This conclusion can be logically valid or invalid. The syllogisms used in Experiment 2 were pretested for the plausibility of their conclusions and for possible floor or ceiling effects (see appendix A for details). The syllogisms have either a logically valid or a logically invalid conclusion. The conclusions are also either plausible or implausible, resulting in *four* different types of conclusions.

Empirical predictions

We assume that people in a neutral state base their conclusion more on the content and the plausibility of a syllogism, whereas people in a state of distrust rely more on the logical consistency of the deduction. However, although the plausibility of a syllogistic conclusion is supposed to have a stronger influence on participants in a neutral state than in a state of distrust, we expect no difference between the two mind-sets as long as the *plausibility* of a conclusion is not itself misleading. Thus, in cases where the conclusions are (1) believable as well as logically valid (congruent), or (2) unbelievable as well as logically invalid (congruent), no difference of the acceptance rate between the two mind-sets should be found. However, we do expect a significant difference in the acceptance rate between the two mind-sets for syllogisms that are (3) believable but logically invalid (incongruent). Under normal conditions, the plausibility of a syllogism should mislead participants to decide that the conclusion is valid, and a belief bias will show up. In a mind-set of distrust, however, where participants rely more on logical considerations, this should happen to a smaller degree.

However, what will happen if a syllogism is (4) unbelievable but logically valid (incongruent)? Unbelievable conclusions should activate analytical thinking and motivate the person to verify the conclusion anew (Evans, 2003; Thompson, Turner, Pennycook, Ball, Brack, Ophir, & Ackermann, 2012). Thus, it is possible that an analytical process will superimpose the heuristic process, which is responsible for the belief bias, and people will therefore recognise the validity of the conclusion better. Because this could happen even in a neutral mind-set, we assume that the performance in a state of distrust will be only slightly better and probably not even significantly better than in the control condition.

Furthermore, we expect participants in a state of distrust to show longer response times, as they are supposed to have more elaborated reasoning during the problem-solving task than participants in a neutral state have.

Methods

Participants and design. Eighty-four students (55 female) with a mean age of M = 22.33 (SD = 3.67) from the University of Bern participated in Experiment 2. Forty-three students received a credit point for their participation, while 41 participated voluntarily. Participants with academic training in logic were excluded from the study in advance. The study was conducted as a 2 x 2 x 2 mixed design, with the validity of the conclusion (valid/invalid) and believability of the conclusion (believable/unbelievable) as within factors, and mind-set (distrust/neutral) as the between factor. We randomly assigned participants to one of the two between conditions. There were 42 participants in each between condition group.

Materials. *Syllogisms.* We used twenty-four pretested syllogisms (see appendix A), translated into German from the syllogisms used by Morley, Evans, & Handley (2004), for the study. Among the 24 syllogisms, 12 had a valid conclusion, and 12 had an invalid conclusion. In addition, half of them featured believable conclusions, and the other half featured unbelievable conclusions, leading to four possible combinations, namely syllogisms with (1) a

believable and valid conclusion, (2) an unbelievable and invalid conclusion, (3) a believable but invalid conclusion, and (4) an unbelievable but valid conclusion.

Procedure. After the participants were welcomed and ushered to their seats in the lab, we informed them of their rights and assured them of the confidentiality of their personal information. They were told that we were interested in obtaining a further understanding of the relationship between language abilities and mathematical abilities. The manipulation for distrust and the neutral condition was realized by employing a subliminal priming procedure (Mayer et al., 2011) carried over from Experiment 1.

Subsequent to the manipulations, participants completed the 12 syllogisms (see appendix for the problems used). The syllogisms (3 per possible combination) were randomly drawn from the pool of 24 pretested syllogisms. For each logical problem, participants had to indicate if the conclusion of the presented syllogism was valid or invalid. There was no time restriction. Upon completion, participants completed the Positive Affect and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988). Finally, the participants answered various demographic questions before we thanked and dismissed them.

Results

Descriptive. We calculated the proportion of accepted conclusions classified as valid for each combination. Table 2 gives an overview of the percentage of acceptance for the different combinations of the syllogisms. The time taken to generate a response can be seen in Table 3.

<< Table 2 >>

<< Table 3 >>

Acceptance rate. A three-way mixed ANOVA revealed a main effect of validity, $F(1,82) = 193.25$, $p < .001$, $\eta_p^2 = .70$, and a main effect of believability, $F(1,82) = 26.14$, $p < .001$, $\eta_p^2 = .24$. No significant main effect of mind-set was found: $F(1,82) = .92$, $p = .34$, $\eta_p^2 = .011$. A significant interaction between believability and validity was revealed: $F(1,82) = 8.87$, $p = .004$, $\eta_p^2 = .10$. Thus, a general belief bias could be demonstrated: across the conditions, valid conclusions were accepted more often than were invalid conclusions (.78 vs. .31), believable conclusions were accepted more often than were unbelievable conclusions (.63 vs. .46), and the effect of believability on accepting conclusions differed according to the validity of the problem.

Planned contrasts revealed, as expected, that there was no difference in the acceptance rate of a conclusion among participants in a distrust mind-set versus a neutral mind-set in cases where the conclusions were unbelievable and invalid (congruent), $t(82) = -1.18$, $p = .24$, $r_{contrast} = .13$, or believable and valid (congruent), $t(82) = 1.43$, $p = .16$, $r_{contrast} = .16$ (conclusions (1) and (2)). As expected, a significant difference between the two mind-set conditions for believable but invalid conclusions (incongruent) occurred: $t(82) = -2.54$, $p = .013$, $r_{contrast} = .27$ (conclusion (3)). In addition, as assumed, the difference in the acceptance rate in cases of unbelievable but valid conclusions was not significant: $t(82) = .80$, $p = .43$, $r_{contrast} = .09$ (conclusion (4)). We found no significant difference in the overall acceptance rate, $t(82) = -1.51$, $p = .14$, $r_{contrast} = .16$.

The results corroborate our assumption that the influence of believability is much weaker for participants under a state of distrust than in a neutral state, especially in those cases where a belief bias is normally most prominent, i.e. for syllogisms with valid but unbelievable conclusions (conclusion (3)) (see Figure 3).

<< Figure 3 >>

Processing times. Overall mean inspection time in seconds amounts to $M = 27.56$ ($SD = 10.12$). A three-way mixed ANOVA showed a main effect of validity, $F(1,82) = 14.49$, $p<.001$, $\eta_p^2 = .15$, and a main effect of believability, $F(1,82) = 8.04$, $p = .006$, $\eta_p^2 = .09$. Participants processed syllogisms with a valid conclusion in a shorter time than syllogisms with an invalid conclusion, and they processed syllogisms with an unbelievable conclusion in a shorter time than they did those with a believable conclusion. As predicted, we found a significant main effect for the between factor mind-set, $F(1,82) = 4.15$, $p = .045$, $\eta_p^2 = .048$. Participants under a state of distrust showed longer processing times than did participants in a neutral state (see Figure 4).

The ANOVA revealed no significant interaction between validity and believability, $F(1,82) = .07$, $p = .78$, $\eta_p^2 < .001$, and no third-order interaction, $F(1,82) = .34$, $p = .56$, $\eta_p^2 = .004$.

<< Figure 4 >>

Negative and positive affect. A one-way ANOVA revealed no significant difference between the two mind-sets concerning positive affect, $F(1,82) = .85$, $p = .358$, $\eta_p^2 = .010$ and negative affect, $F(1,82) = .14$, $p = .714$, $\eta_p^2 = .002$; using the PANAS (α for Positive Affect Scale = .81, α for Negative Affect Scale = .82).

Discussion

Participants in the distrust condition and in the control condition did not differ in their acceptance rates of a syllogistic conclusion as long as the believability and validity of the conclusions were congruent, i.e., either believable and valid or unbelievable and invalid (conclusions (1) and (2)). We expected this result because in those cases, the belief bias does not mislead the conclusion of participants. Again, we found no significant difference of the acceptance rate for the two mind-sets if conclusions were unbelievable but invalid (conclusion (4)). This corroborates findings of other studies that unbelievable conclusions normally

trigger a more elaborated reasoning process and that this elaboration increases the possibility of finally detecting the validity of the conclusion (Evans, 2003; Thompson et al., 2012). According to the parallel dual process models (Evans, 2006, 2008; Stanovich, 2004), unbelievable conclusions activate analytical thinking and motivate the logical analysis of arguments; thus, in these cases, it is possible to superimpose the heuristic process that is associated with the belief bias. Thus, if the unbelievabilty of a conclusion enhances the detection of a valid conclusion in general, the additional effect of distrust may decline to an insignificant quantity, as we could show.

However, for invalid but believable conclusions (conclusion (3)), distrustful participants showed a significantly lower acceptance rate than participants in a neutral state. The latter showed the well-documented degree of fallibility of human reasoning if the syllogisms are believable but actually invalid. This result was also demonstrated in our pre-study (Appendix B). Contrary to a neutral state, distrust leads, in this case, to a significant reduction of the belief bias. Participants in a state of distrust seem to focus mainly on the logical validity of a conclusion and, as a result, are less mislead by the believability of a conclusion. They assess the validity of the most problematic syllogisms extraordinarily well.

In addition, under a state of distrust (vs. neutral state), participants take significantly more time to come to a decision. This supplements our assumption that distrust leads to a more elaborated reasoning while participants in a neutral state rely more on quick, heuristic judgments, such as “if a conclusion is believable, it is also logically correct”.

General Discussion

Being in a state of distrust has positive functions; e.g., it turns people to non-routine thought processes (Schul et al., 2008, Mayer et al., 2011). However, from an evolutionary point of view, distrust should especially serve the function of detecting cheating or lying sooner and more accurately. Unfortunately, this question has been scarcely investigated. One of these few investigations is the study by Miller et al. (1986). They examined directly

whether distrust improves deception detection. Moreover, Fein (1996) investigated whether distrust facilitates the detection of an adopted opinion, not the own one.

In Experiment 1, we corroborated the finding of Miller et al. (1986): distrust improves the detection of a falsified report without declining the correct detection of the true report. This finding is important because it falsifies the probably obvious assumption that distrust only increases the tendency to indicate that someone is lying (lie bias) but does not improve the accuracy of distinguishing true from falsified reports. If a lie bias would be responsible for the improvement in detecting lies under distrust, then it would do so only at the expense of correctly detecting the truth. Our data show clearly that this is not the case.

The qualitative analysis of participants' specified cues that may have guided their veracity judgments gives us a first indication of why distrust improved the detection of lies without worsening the truth detection. Participants in a state of distrust report that they mainly employ logical inconsistency while participants in a neutral state report that they pay more attention to the plausibility of statements. If self-reports could be considered as valid, the improvement of lie detection under distrust could be explained as follows: Because logical consistency is a more valid cue of lie detection than plausibility/believability of the statements, distrustful participants have better chances to discriminate between falsified and true reports than those in a neutral state. Unfortunately, self-reports are not necessarily valid, however. To examine the explanation with more rigour, it would be necessary to have several reports about incidents where reality status (true/falsified) and the two features (logical consistency/plausibility) vary independently. Because logical consistency is normally confounded with the reality status in ecologically valid statements, it would be difficult to find comparable reports that fulfil this condition, however.

Therefore, in Experiment 2, we changed the paradigm. Regarding our assumption, distrust should prevent people from erroneously using plausibility/believability as an assumed valid cue in their judgments. Therefore, their belief bias should be less pronounced than for

people in a neutral state. The belief bias means that believability erroneously influences a judgment. An erroneous influence of believability can only be the case, however, if this judgment is at the same time logically invalid or false because of other reasons. Especially in this case, people will accept the judgment the more as correct or valid the higher their belief bias is pronounced.

In Experiment 2, we used syllogisms to examine whether distrust decreases the belief bias. Four different kinds of syllogisms can be differentiated: syllogisms with a (1) valid and believable, (2) valid and unbelievable, (3) invalid and believable, (4) invalid and unbelievable conclusion. The first two kinds of syllogism are unimportant for the examination of a belief bias because believability cannot mislead participants if both features are congruent. The fourth kind of syllogism is of minor importance because it is assumed that the implausibility of a conclusion triggers a higher elaboration of reasoning even in normal conditions. Thus, differences in the belief bias between participants in a state of distrust versus a neutral state should be small. Most important is the third kind of syllogism, however, where believable conclusions are logically *invalid*. Here, we expected that distrustful participants would show a significantly smaller belief bias than the other participants. Experiment 2 support these assumptions. We found no difference between the distrust and control conditions in the acceptance rate for the first two kinds of syllogisms with congruent conclusions. Thus, we could show that distrust had no advantage if logic and plausibility are congruent, that is, if a syllogism was either valid and believable or invalid and unbelievable. In those cases, a belief bias should not prevent a correct solution, and thus, participants in one condition were as good as in the other. However, we could prove that under distrust, participants seem to be less prone to a belief bias. There was a small but insignificant and less pronounced belief bias among distrustful participants if they had to solve the fourth kind of syllogism with unbelievable but valid conclusions, and a significantly lower belief bias of distrustful participants if they had to solve the third, and central kind of syllogism, with believable but

invalid conclusions. Additional support for our assumption that distrust goes along with a more elaborated reasoning process is corroborated by the finding that participants in the distrust condition take more time to proceed and to come to a judgment.

As supposed, distrust may enhance analytical thinking and prevents the usage of a more accessible cue, such as believability. Under distrust, the belief bias is reduced, indicating that participants started to think analytically about the logical structure of the syllogisms.

According to parallel dual process models, unbelievable conclusions activate analytical thinking and motivate the logical analysis of arguments; thus, it is possible to superimpose the heuristic process that is associated with the belief bias (Evans, 2006, 2008; Stanovich, 2004). People under distrust rely ab initio more on logical structure as a basis for their judgment. Thus, falsified reports should be recognized as such, regardless of whether they seem plausible or not.

Thus, people under distrust show greater cognitive flexibility. Schul et al. (2008) showed that participants primed with distrust were sensitive to non-routine contingencies and made use of them more often. In line with this, Mayer et al. (2011) put forward that people under a state of distrust become more creative through the process of cognitive flexibility. Our study adds more insight to this field of research. We showed that distrust also enhances the correct detection of lies and, furthermore, a correct classification of conclusions as valid or invalid. Further research should investigate the basic cognitive processes that are responsible for all the manifestations. Is it cognitive control? As Groborz & Necka (2003) point out in two experiments, cognitive control is associated with idea production and a more accurate evaluation of the ideas of others. Additionally, Benedek, Franz, Heene, & Neubauer (2012) pointed out a positive correlation of inhibition and creativity measures. Subliminally induced distrust constitutes an automatic activation of cognitive control (Lau & Passingham, 2007; van Gaal, Ridderinkhof, Scholte, & Lamme, 2010) in the prefrontal cortex. Further studies in this field are necessary and will bring a deeper insight into the processes that trigger the

logical system in a state of distrust. Furthermore, the occurrence, the relevance, and the consequences of determining an unbelievable statement as invalid in a neutral state should be investigated in detail in a more applied area.

Limitations

Even though we could mainly corroborate our theoretical assumptions, we have to mention some limitations of the experiments. Unfortunately, we selected in Experiment 1 a true ort report of an incident that had a high a priori rate of being judged correctly (about 90%). Additionally, one could criticise that we should have used more than just two reports in order to gain data that are more reliable and more meaningful. This may be improved in further studies, if the difficulty can be overridden to find enough reports that are comparable in most respects but where reality status (true/false) and the relevant features (logical consistency/ plausibility) vary independently. Another limitation exists because we manipulated distrust and the neutral state of mind only with one method. We preferred to use an implicit method, such as a subliminal priming method, because the implicit manipulation proved to be successful. Furthermore, we were afraid that an explicit manipulation of distrust could trigger unwanted control processes (Strahan, Spencer, & Zanna, 2002).

Experiment 2 also has some limitations. We only recorded the reaction times as variables to depict the process during the judgment phase. Gaining data from introspection or neurophysiological parameters would be interesting for further studies. Goel, Buchel, Frith, and Dolan (2000) have already detected neuronal correlates of the belief bias. They speculate that the right prefrontal cortex is involved when it comes to generating a correct answer, because it is a critical area for detecting a conflict between belief and logic, as well as for resolving one. The activation of the ventromedial prefrontal cortex, however, emphasizes its role in belief-based responses, which might lead to errors. In further studies, the insights of rather psychological, cognitive scientific studies on one hand and neuroscientific studies on the other hand should be combined.

Concerning the processing times, our results show that participants took less time to think about unbelievable conclusions. Thompson et al. (2003) also showed this, at first rather surprising, result: “Reasonless spent less time reasoning about unbelievable conclusions than about believable ones”. Although some studies have shown that people need more time to analyse an unbelievable conclusion logically (Evans, Newstead, Allen, & Pollard, 1994; Newstead, Pollard, Evans, & Allen, 1992), the study by Thompson et al. (2003) clearly shows that people needed more time to analyse a credible conclusion. It is of interest to find out which variables are responsible for the contradictory results.

Conclusion

As research revealed, distrust has positive effects on detecting deception and exposing liars (Miller et al., 1986; Hilton, Fein, & Miller, 1993; Fein, 1996). People in a state of distrust become better at detecting a falsified report without ignoring the truth. When asked which cues they used to evaluate the veracity of a report, subjects seem to rely on a logic-based analytical reasoning process. This rather subjective and qualitative measurement was underpinned by an objective, quantitative measurement. As we found out, people in a state of distrust clearly showed a reduced belief bias. This further indicates that they used an analytical, logical thought process when evaluating a conclusion as valid or invalid. A state of distrust may trigger a bottom-up thought process, probably inhibiting the prepotent response, and therefore functions as a cognitive control.

References

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology, 10,* 461-471. doi:10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences*, *53*, 480-485. doi: 10.1016/j.paid.2012.04.014

Bies, R.J., & Tripp, T.M. (1996). Beyond distrust: "Getting even" and the need for revenge. In R.M. Kramer and T. Tyler (Eds.), *Trust in organizations* (pp. 246-260). Newbury Park: Sage Publications.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214-234.doi: 10.1207/s15327957pspr1003_2

Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In Ashmore, R., Jussim, L., & Wilder, D. (Eds). *Social identity, intergroup conflict, and conflict reduction*. New York: Oxford University Press.

Chambers, J. R., & Melnyk, D. (2006). Why do I hate thee? Conflict misperceptions and intergroup mistrust. *Personality and Social Psychology Bulletin, 32*, 1295-1311. doi: 10.1177/0146167206289979

Cho, J. (2006). The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing, 82,* 25-35. doi: 10.1016/j.jretai.2005.11.002

DePaulo, B. M. & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*, 63-79. doi:10.1037/0022-3514.74.1.63

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979 - 995. doi:10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-118. doi:10.1037/0033-2909.129.1.74

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations, 13*, 123-139. doi: 10.1177/001872676001300202

Diener, E., & Seligman, M. E. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*, 1-31. doi: 10.1111/j.0963-7214.2004.00501001.x

Evans, J. St. B. T (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences,7*, 454-459. doi:10.1016/j.tics.2003.08.012

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review,13*, 378-395. doi:10.3758/BF03193858

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11*, 295-306. doi: 10.3758/BF03196976

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 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*, 112-119. doi:10.1177/1948550610382666

Früh, W. (2011). *Inhaltsanalyse: Theorie und Praxis* (Vol. 2501). UTB.

Gigerenzer, G. (2008). *Gut Feelings: Short Cuts to Better Decision Making.* London: Penguin Books Limited.

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

Goel, V. (2003). Evidence for dual neural pathways for syllogistic reasoning. *Psychologica*, *32*, 301-309.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage, 12,* 504-514. doi. 10.1006/nimg.2000.0636

Groborz, M., & Necka, E. (2003). Creative and cognitive control: Explorations of generation and evaluation skill. *Creativity Research Journal, 14*, 55-67.doi: 10.1016/j.paid.2012.04.014Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*. doi: 10.1002/acp.3052

Hill, C., Memon, A. & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology, 13*, 357-371. doi:10.1348/135532507X238682

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin, 19,* 501-512. doi:10.1177/0146167293195003

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior, 27*, 187-203. doi:10.1023/A:1022599230598

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107*, 852-884. doi:10.1037/0033-295X.107.4.852

Kramer, R. M., Brewer, M. B, & Hanna, B. (1996). Collective trust and collective action in organizations: The decision to trust as a social decision. In R. M. Kramer and T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 357-289). Thousand Oaks, CA: Sage.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience, 27*,5805-5811. doi:10.1523/JNEUROSCI.4335-06.2007

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review, 23*, 438-458.doi: 10.5465/AMR.1998.926620

March, J. G., & Olsen, J. P. (1975). The uncertainty of the past: Organizational learning under ambiguity*. *European Journal of Political Research*, *3*(2), 147-171. doi: 10.1111/j.1475-6765.1975.tb00521.x

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709-734.doi: 10.5465/AMR.1995.9508080335

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

Mayring, P. (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken (10th edition).*Weinheim: Beltz.

Miller, G. R., Mongeau, P. A., & Sleight, C. (1986). Fudging with friends and lying to lovers: Deceptive communication in personal relationships. *Journal of Social and Personal Relationships, 7.* 107-118. doi:10.1177/0265407586034006

Morgan, J. J. & Morton, J. T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *The Journal of Social Psychology, 20*, 39-59. doi:10.1080/00224545.1944.9918830.

Morley, N. J., Evans, J. S. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57*, 666-692. doi:10.1080/02724980343000440

Park, H. S., Levine, T., McCornack, S., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs, 69,* 144-157. doi:10.1080/714041710

Posten, A. -C. & Mussweiler, T. (2011). Priming trust. Manuscript in preparation.

Posten, A.-C., & Mussweiler, T. (2013). When distrust frees your mind: The stereotype-reducing effects of distrust. *Journal of Personality and Social Psychology, 105*, 567–584. doi:10.1037/a0033170

Reinhard, M. A., & Schwarz, N. (2012). The influence of affective states on the process of lie detection. *Journal of Experimental Psychology: Applied, 18*, 377-389. doi: 10.1037/a0030466.

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*, 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*, 1293-1302. doi:10.1016/j.jesp.2008.05.003

Sitkin, S.B.,& George, E. 2005. Managerial trust-building through the use of legitimating formal and informal control mechanisms. *International Sociology, 20*, 307-338.doi: 10.1177/0268580905055479

Sitkin, S. B. & Stickel, D. (1996). The road to hell: The dynamics of distrust in an era of quality. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations. Frontiers in theory and research* (pp. 196-216*).* Thousand Oaks, CA: Sage Publications.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

Steinbach, J. (2011). *Erkennung von Täuschung in schriftlichen Aussagen Angeschuldigter, welche sich eines Fahrlässigkeitsdelikts schuldig gemacht haben. [Detecting deception in written statements of suspects who are guilty of negligence]*. Unpublished manuscript.

Steller, M. & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods for investigation and evidence* (pp. 217-245). New York: Springer.

Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology, 38,* 556-568.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*, 237-251doi: 10.1016/j.cognition.2012.09.012.

van Gaal, S., Ridderinkhof, K. R., Scholte, H. S., & Lamme, V. A. (2010). Unconscious activation of the prefrontal no-go network. *The Journal of Neuroscience, 30*, 4143-4150. doi:10.1523/JNEUROSCI.2992-09.2010

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: Wiley-Interscience.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: the role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29,* 141-208. doi: 10.1016/S0065-2601(08)60017-9

Wilkins, M. C. (1929). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology, 102.*

Appendix A

It was the aim of this pretest to check whether the subliminal priming of distrust (versus control) is successful. Thus, the pretest presents a manipulation check.

Method and Procedures

Participants and design. Forty students from the University of Bern participated voluntarily in the pretest. Participants were randomly assigned to one of two conditions, namely a distrust and a neutral condition. In the distrust condition (n = 20) mean age was M = 30.15 (SD=10.32) and in the neutral condition (n =20) mean age was M = 28.14 (SD = 12.25).

The independent variables were identical for all participants. We checked the influence of the subliminal induction of distrust on the explicitly measured variables trust, negative affect, positive affect and feelings of trust and distrust. Additionally we recorded trust and distrust as well as negative and positive affect in an implicit way.

Material. ***Subliminal Priming.*** The subliminal priming, which was used to induce distrust manipulation, was programmed with the Computer Software Inquisit 3.0.5.0 (Millisecond, 2011). The priming comprised 2 block, 24 trials each, with two additional exercise trials at the beginning of each block. In the distrust conduction the word „misstrauen“ served as prime, in the neutral condition no prime was shown at all. The subliminal prime was presented for 13 ms; it was follow by a mask for 506 ms until the target appeared. The target was presented for 3000 ms, followed by a mask, whereupon the next prime appeared. The target was either a word or a nonword. The chronology of the subliminal priming corresponds to the recommendations of Bargh and Chartrand (2000). Participants had the task to decide if a target was a word or a nonword. The subliminal priming was designed as a lexical decision task.

Manipulation check (independent variables). Positive and negative affect, trust and distrust were measured implicitly (IPANAT; Quirin, Kazén, & Kuhl, 2009). The IPANAT measures affect in an indirect way by word-associations. Participants judged six artificial

words (SAFME, VIKES, TUNBA, TALEP, BELNI, SUKOV) concerning six attributes on a 4-point scale (1 = does not fit at all, 2 = fits somewhat, 3 = fits quite well, 4 = fits very well). The attributes for the measurement of implicit negative and positive affect were: (happy [fröhlich], cheerful [gut gelaunt], energetic [aktiv]), (helpless [hilflos], tense [verkrampft], inhibited [gehemmt]). For the measurement of implicit distrust and trust the attributes were: (distrustful [misstrauisch], deceitful [hinterlistig], suspicious [verdächtig], sincere [aufrichtig], trustworthy [glaubwürdig]. Additionally we measured trust (questionnaire with ten items; Dunn & Schweitzer, 2005) as well as positive and negative affect (PANAS; Watson & Tellegen, 1988) explicitly. Feelings of trust and distrust were measured with one item respectively.

Procedure. The entire pretest was computer based. Following the subliminal priming, participants processed the implicit measures in random order and subsequently the explicit measures, again in random order. Finally, participants had to indicate their current feelings of trust and distrust and demographics were collected.

Results and Discussion

Table A-1 gives an overview over the dependent variables.

Tabelle A-1

Mean (M), Standard Deviation (SD) of the dependent variables sorted by condition (incl. Cronbach's alpha)

Condition subliminal priming		Distrust $n = 20$	Neutral $n = 20$
	Cronbach's Alpha	*M* (*SD*)	*M* (*SD*)
Implicit trust (IPANAT)	α = .65	2.29 (.34)	2.27 (.28)
Implicit distrust (IPANAT)	α = .80	2.14 (.49)	1.83 (.39)
Implicit positive affect (IPANAT)	α = .60	2.26 (.37)	2.27 (.34)
Implicit negative affect (IPANAT)	α = .74	1.83 (.38)	1.88 (.38)

Explicit trust (Dunn et al., 2005).	α = .86	32.65 (7.67)	35.30 (6.03)
Explicit positive affect (PANAS)	α = .80	30.40 (4.55)	32.80 (6.35)
Explicit negative affect (PANAS)	α = .90	15.10 (6.83)	13.45 (6.01)
Feeling of trust	1 item	73.95 (22.84)	73.95 (17.31)
Feeling of distrust	1 item	38.80 (30.67)	25.25 (23.35)

Implicit measures. Implicit negative affect and positive affect did not significantly differ between the conditions, $t(38)=.441$, $p=.662$, resp. $t(38)=.122$, $p=.904$. However, differences between the conditions were found for implicit distrust $t(38) = 2.246$, $p=.031$, but not for implicit trust, $t(38) = .139$, $p=.890$.

Explicit measures. We found no significant difference with the trust-questionnaire between the two conditions, $t(38) = -1.215$, $p = .165$, nor did we find any significant difference between the two conditions for explicit negative affect, $t(38) = .811$, $p = .211$, explicit positive affect, $t(38) = -1.374$, $p = .089$ and feeling of trust, $t(38) = 0$, $p = 1.00$. For the feeling of distrust we could reveal a strong tendency, $t(38) = 1.572$, $p = .062$.

Discussion. The pretest shows, that inducing distrust with a subliminal priming paradigm is possible. The results of the implicit measures corroborate the expected difference between the distrust and the neutral condition. The subliminal priming of distrust raises implicit distrust but does not increase a negative affect simultaneously. Moreover, participants in the distrust condition seem to have a stronger feeling of distrust than participants in the neutral condition. The results of the pretest are satisfactory and designate the subliminal priming paradigm as valid possibility to induce a state of distrust. Even with a trust-questionnaire we could find a small but insignificant difference between the two mind-sets in the expected direction. Import seems to be the finding that the induction of distrust has only a negligible influence on affect (implicitly and explicitly measured). Thus, it is possible to distinguish distrust from a negative affect.

Appendix B

The original syllogisms by Morley, Evans, & Handley (2004) were translated to German and pretested. The syllogisms contained in the premises only "no" (E) and "some" (I) or "some" (I) and "no" (E) and in the conclusion always "some... not" (O). Twelve syllogisms were in the form EIO and 12 in the form IEO. Half of the syllogisms presented were valid and half invalid. Half of the syllogisms used had an implausible conclusion and the other half had a credible conclusion. Two figures were used: XY-YZ and XY-ZY (Johnson-Laird, 1991).

Believability of the Conclusions

A priori, as believable categorized conclusions were evaluated as more believable. (M = 5.59, SD = .40) as those categorizes as unbelievable (M = 2.02, SD = .43), $t(42) = 33.10$, $p < .001$, $d_z = 5.05$. See table B-1 for details.

Table B-1

Believability rating of the conclusions of the syllogisms used in experiment 2.

conclusion	*M*	*SD*
believable		
Einige gute Schwimmer sind keine Tiefseetaucher.	5.72	.77
Einige religiöse Menschen sind keine Priester.	5.67	.75
Einige Vögel sind keine Sperlinge.	5.37	1.27
Einige suchterzeugende Dinge sind keine Zigaretten.	5.63	.90
Einige gut ausgebildete Menschen sind keine Juristen.	5.72	.59
Einige Tiere sind keine Katzen.	5.56	1.33
Einige gesunde Menschen sind keine Astronauten.	5.65	1.29
Einige gute Schwimmer sind keine Tiefseetaucher.	5.86	.35
Einige Geflügel sind keine Hühner.	4.51	1.94
Einige gut trainierte Hunde sind keine Polizeihunde.	5.63	1.09
Einige gesunde Menschen sind keine Astronauten.	5.98	.15
Einige Fische sind keine Forellen.	5.93	.26
	5.59	.40

unbelievable		
Einige Priester sind nicht religiös.	2.53	1.32
Einige Millionäre sind nicht reich.	2.12	1.61
Einige Cockerspaniel sind keine Hunde.	1.23	.53
Einige Millionäre sind nicht reich.	2.44	1.79
Einige Zigaretten sind nicht suchterzeugend.	1.63	1.07
Einige Schlangen sind keine Reptilien.	1.26	.79
Einige Richter sind nicht gut ausgebildet.	2.65	.95
Einige Polizeihunde sind nicht gut trainiert.	2.28	1.24
Einige Wale sind keine Säugetiere.	1.65	1.23
Einige Vitamine sind keine Nahrungsmittel.	3.30	1.85
Einige Lebensmittel sind nicht essbar.	1.84	1.41
Einige Käfer sind keine Insekten.	1.77	1.31
	2.02	.43

Belief-Bias

An overview of the acceptance rates, or the rate of subjects who considered a conclusion as valid, is given in Table B-2. A 2 x 2 ANOVA with repeated measures indicates a significant main effect of believability of the conclusion, $F(1, 42) = 61.02$, $p < .001$, $\eta p2 = .59$, and a significant main effect of validity of the conclusion, $F(1, 42) = 34.79$, $p < .001$, $\eta p2 = .45$.

Table B-2

Acceptance rates Mean M and (Standard Deviation SD)

	believable	unbelievable	
valid	.77 (.26)	.57 (.35)	.67 (.23)
invalid	.67 (.30)	.19 (.30)	.43 (.22)
	.72 (.23)	.37 (.24)	

Analogous to Evans et al. (1983), we conclude that believable conclusions (72%) are more likely to be accepted than unbelievable (37%). However, subjects definitely have logical competence as they accept valid conclusions (67%) to a greater extent than invalid

ones (43%). In addition, the credibility x validity interaction was significant: $F(1, 42) = 11.61$, $p = .001$, $\eta_p^2 = .22$. The belief bias is more pronounced for invalid than for valid conclusions: $t(42) = 7.86$, $p < .001$, $d_z = .51$ vs., $t(42) = 3.37$, $p < .001$, $d_z = 1.20$.

Table B-3

Syllogisms used in Experiment 2

	form	
combination	EIO	IOE
valid- believable	Keine guten Schwimmer sind Pendler. Einige Pendler sind Tiefseetaucher. Folgt: Einige gute Schwimmer sind keine Tiefseetaucher. (figure XY-YZ) Keine Tiefseetaucher sind Raucher. Einige Raucher sind gute Schwimmer. Folgt: Einige gute Schwimmer sind keine Tiefseetaucher. (figure XY-ZY) Keine Hühner sind Junarics. Einige Junarics sind Geflügel. Folgt: Einige Geflügel sind keine Hühner. (figure XY-ZY)	Einige gesunde Menschen sind unglücklich. Keine unglücklichen Menschen sind Astronauten. Folgt: Einige gesunde Menschen sind keine Astronauten. (figure XY-YZ) Einige Vögel sind Phylone Keine Phylone sind Sperlinge. Folgt: Einige Vögel sind keine Sperlinge. (figure XY-YZ) Einige Priester sind Gasableser. Keine Gasableser sind religiöse Menschen. Folgt: Einige religiöse Menschen sind keine Priester (figure XY-ZY)
valid- unbelievable	Keine religiösen Menschen sind gesund. Einige gesunde Menschen sind Priester. Folgt: Einige Priester sind nicht religiös. (figure XY-ZY) Keine gut trainierten Hunde sind Raben. Einige Raben sind Polizeihunde.	Einige Richter sind faul. Keine faulen Menschen sind gut ausgebildet. Folgt: Einige Richter sind nicht gut ausgebildet. (figure XY-YZ) Einige Millionäre sind Schwätzer. Keine Schwätzer sind reich.

	Folgt: Einige Polizeihunde sind nicht gut trainiert. (figure XY-ZY) Keine Säugetiere sind Enkulions. Einige Enkulions sind Wale. Folgt: Einige Wale sind keine Säugetiere. (figure XY-ZY)	Folgt: Einige Millionäre sind nicht reich. (Figur XY-YZ) Einige Cockerspaniel sind Biktoide. Keine Biktoide sind Hunde. Folgt: Einige Cockerspaniel sind keine Hunde. (figure XY-YZ)
invalid-believable	Keine gut trainierten Hunde sind bösartig. Einige bösartigen Hunde sind Polizeihunde. Folgt: Einige gut trainierte Hunde sind keine Polizeihunde. (figure XY-YZ) Keine gut ausgebildeten Menschen sind Monteure. Einige Monteure sind Juristen. Folgt: Einige gut ausgebildete Menschen sind keine Juristen. (figure XY-YZ) Keine Tiere sind Kryptoden. Einige Krypotden sind Katzen. Folgt: Einige Tiere sind keine Katzen. (figure XY-YZ)	Einige Zigaretten sind preiswert. Keine preiswerten Dinge sind suchterzeugend. Folgt: Einige suchterzeugende Dinge sind keine Zigaretten. (figure XY-ZY) Einige Astronauten sind Fischer. Keine Fischer sind gesund. Folgt: Einige gesunde Menschen sind keine Astronauten. (figure XY-ZY) Einige Forellen sind Hämophede. Keine Hämophede sind Fische. Folgt: Einige Fische sind keine Forellen. (figure XY-ZY)
invalid-unbelievable	Keine Millionäre sind fleissige Arbeiter. Einige fleissige Arbeiter sind reich. Folgt: Einige Millionäre sind nicht reich. (figure XY-YZ) Keine Lebensmittel sind kandiert Einige kandierte Dinge sind essbar.	Einige Nahrungsmittel sind preiswert. Keine preiswerten Dinge sind Vitamine. Folgt: Einige Vitamine sind keine Nahrungsmittel. (figure XY-ZY) Einige suchterzeugende Dinge sind alkoholhaltig. Keine alkoholhaltigen Dinge sind Zigaretten.

	Folgt: Einige Lebensmittel sind nicht essbar. (figure XY-YZ) Keine Käfer sind Zaphoden. Einige Zaphoden sind Insekten. Folgt: Einige Käfer sind keine Insekten. (figure XY-YZ)	Folgt: Einige Zigaretten sind nicht suchterzeugend. (figure XY-ZY) Einige Reptilien sind Glissome. Keine Glissome sind Schlangen. Folgt. Einige Schlangen sind keine Reptilien. (figure XY-ZY)

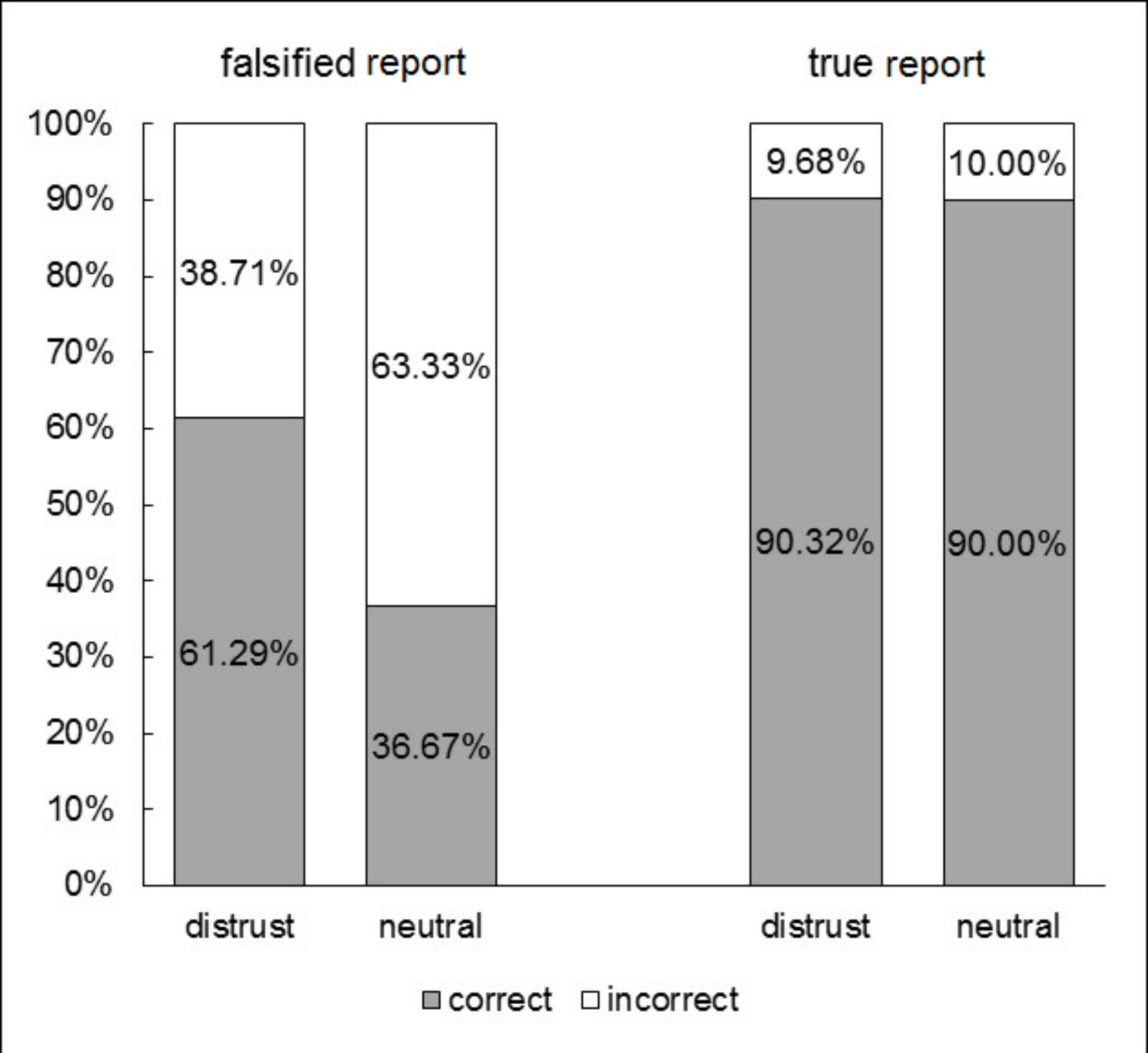
Table B-3

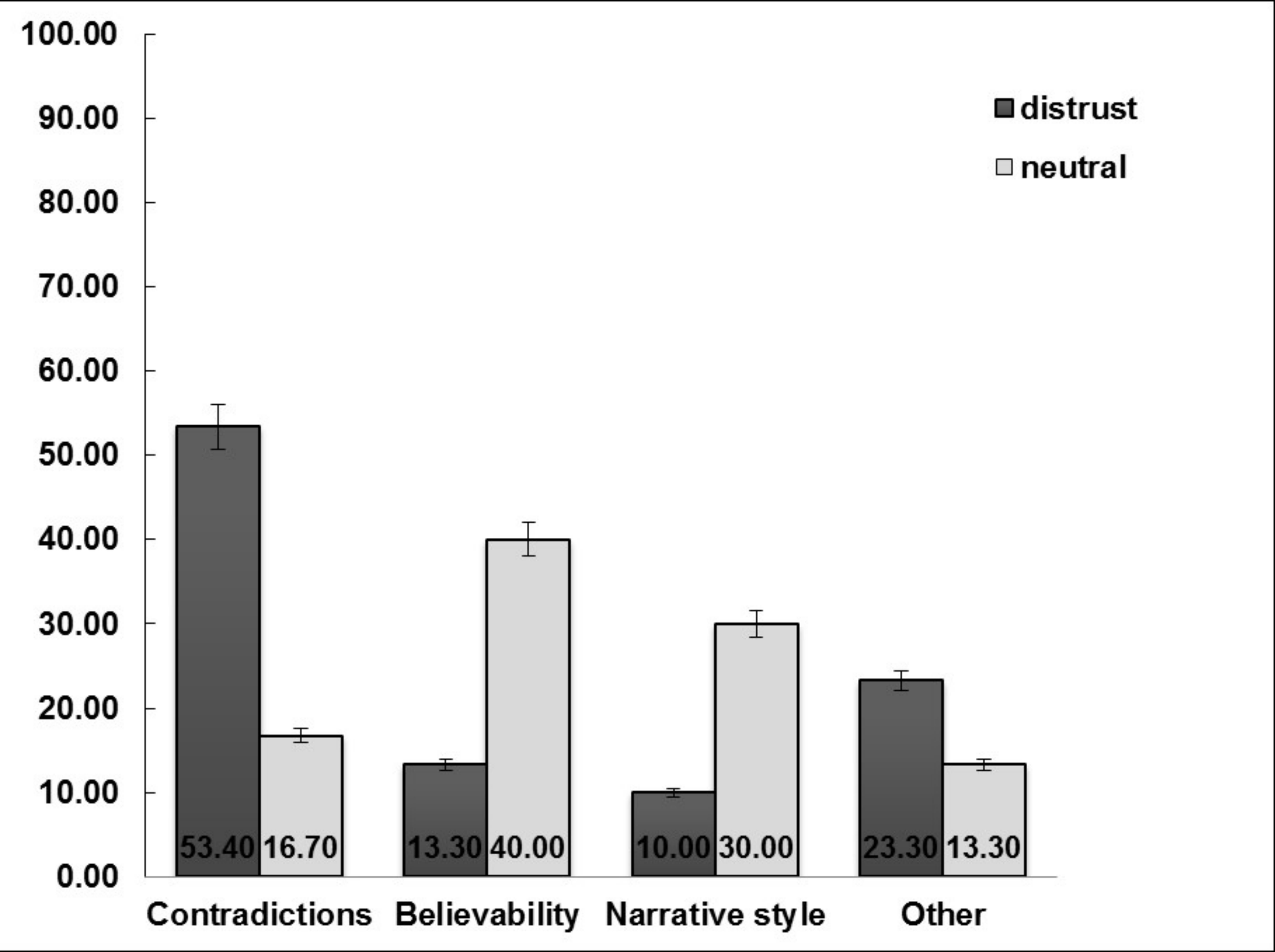
Syllogisms used in Experiment 2

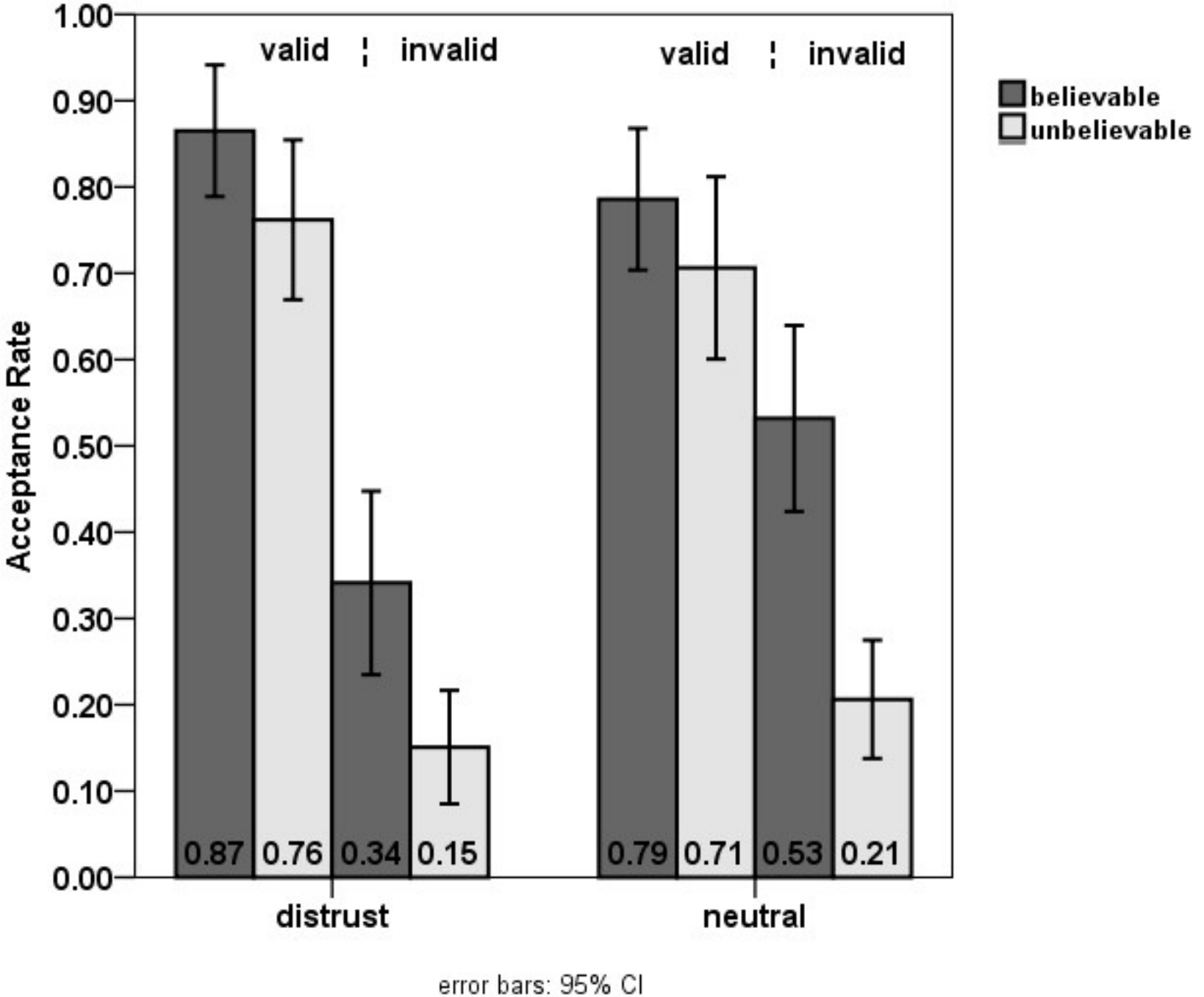
	form	
combination	EIO	IOE
valid- believable	Keine guten Schwimmer sind Pendler. Einige Pendler sind Tiefseetaucher. Folgt: Einige gute Schwimmer sind keine Tiefseetaucher. (figure XY-YZ) Keine Tiefseetaucher sind Raucher. Einige Raucher sind gute Schwimmer. Folgt: Einige gute Schwimmer sind keine Tiefseetaucher. (figure XY-ZY) Keine Hühner sind Junarics. Einige Junarics sind Geflügel. Folgt: Einige Geflügel sind keine Hühner.	Einige gesunde Menschen sind unglücklich. Keine unglücklichen Menschen sind Astronauten. Folgt: Einige gesunde Menschen sind keine Astronauten. (figure XY-YZ) Einige Vögel sind Phylone Keine Phylone sind Sperlinge. Folgt: Einige Vögel sind keine Sperlinge. (figure XY-YZ) Einige Priester sind Gasableser. Keine Gasableser sind religiöse Menschen. Folgt: Einige religiöse Menschen sind keine Priester

	(figure XY-ZY)	(figure XY-ZY)
valid-unbelievable	Keine religiösen Menschen sind gesund. Einige gesunde Menschen sind Priester. Folgt: Einige Priester sind nicht religiös. (figure XY-ZY) Keine gut trainierten Hunde sind Raben. Einige Raben sind Polizeihunde. Folgt: Einige Polizeihunde sind nicht gut trainiert. (figure XY-ZY) Keine Säugetiere sind Enkulions. Einige Enkulions sind Wale. Folgt: Einige Wale sind keine Säugetiere. (figure XY-ZY)	Einige Richter sind faul. Keine faulen Menschen sind gut ausgebildet. Folgt: Einige Richter sind nicht gut ausgebildet. (figure XY-YZ) Einige Millionäre sind Schwätzer. Keine Schwätzer sind reich. Folgt: Einige Millionäre sind nicht reich. (Figur XY-YZ) Einige Cockerspaniel sind Biktoide. Keine Biktoide sind Hunde. Folgt: Einige Cockerspaniel sind keine Hunde. (figure XY-YZ)
invalid-believable	Keine gut trainierten Hunde sind bösartig. Einige bösartigen Hunde sind Polizeihunde. Folgt: Einige gut trainierte Hunde sind keine Polizeihunde. (figure XY-YZ) Keine gut ausgebildeten Menschen sind Monteure. Einige Monteure sind Juristen. Folgt: Einige gut ausgebildete Menschen sind keine Juristen. (figure XY-YZ) Keine Tiere sind Kryptoden. Einige Krypotden sind Katzen. Folgt: Einige Tiere sind keine Katzen.	Einige Zigaretten sind preiswert. Keine preiswerten Dinge sind suchterzeugend. Folgt: Einige suchterzeugende Dinge sind keine Zigaretten. (figure XY-ZY) Einige Astronauten sind Fischer. Keine Fischer sind gesund. Folgt: Einige gesunde Menschen sind keine Astronauten. (figure XY-ZY) Einige Forellen sind Hämophede. Keine Hämophede sind Fische. Folgt: Einige Fische sind keine Forellen.

	(figure XY-YZ)	(figure XY-ZY)
invalid-unbelievable	Keine Millionäre sind fleissige Arbeiter. Einige fleissige Arbeiter sind reich. Folgt: Einige Millionäre sind nicht reich. (figure XY-YZ) Keine Lebensmittel sind kandiert Einige kandierte Dinge sind essbar. Folgt: Einige Lebensmittel sind nicht essbar. (figure XY-YZ) Keine Käfer sind Zaphoden. Einige Zaphoden sind Insekten. Folgt: Einige Käfer sind keine Insekten. (figure XY-YZ)	Einige Nahrungsmittel sind preiswert. Keine preiswerten Dinge sind Vitamine. Folgt: Einige Vitamine sind keine Nahrungsmittel. (figure XY-ZY) Einige suchterzeugende Dinge sind alkoholhaltig. Keine alkoholhaltigen Dinge sind Zigaretten. Folgt: Einige Zigaretten sind nicht suchterzeugend. (figure XY-ZY) Einige Reptilien sind Glissome. Keine Glissome sind Schlangen. Folgt. Einige Schlangen sind keine Reptilien. (figure XY-ZY)


falsified report
true report
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
38.71%
61.29%
63.33%
36.67%
9.68%
90.32%
10.00%
90.00%
distrust
neutral
distrust
neutral
correct
incorrect


100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
distrust
neutral
53.40
16.70
13.30
40.00
10.00
30.00
23.30
13.30
Contradictions
Believability
Narrative style
Other


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Acceptance Rate
valid
invalid
valid
invalid
believable
unbelievable
0.87
0.76
0.34
0.15
0.79
0.71
0.53
0.21
distrust
neutral
error bars: 95% CI

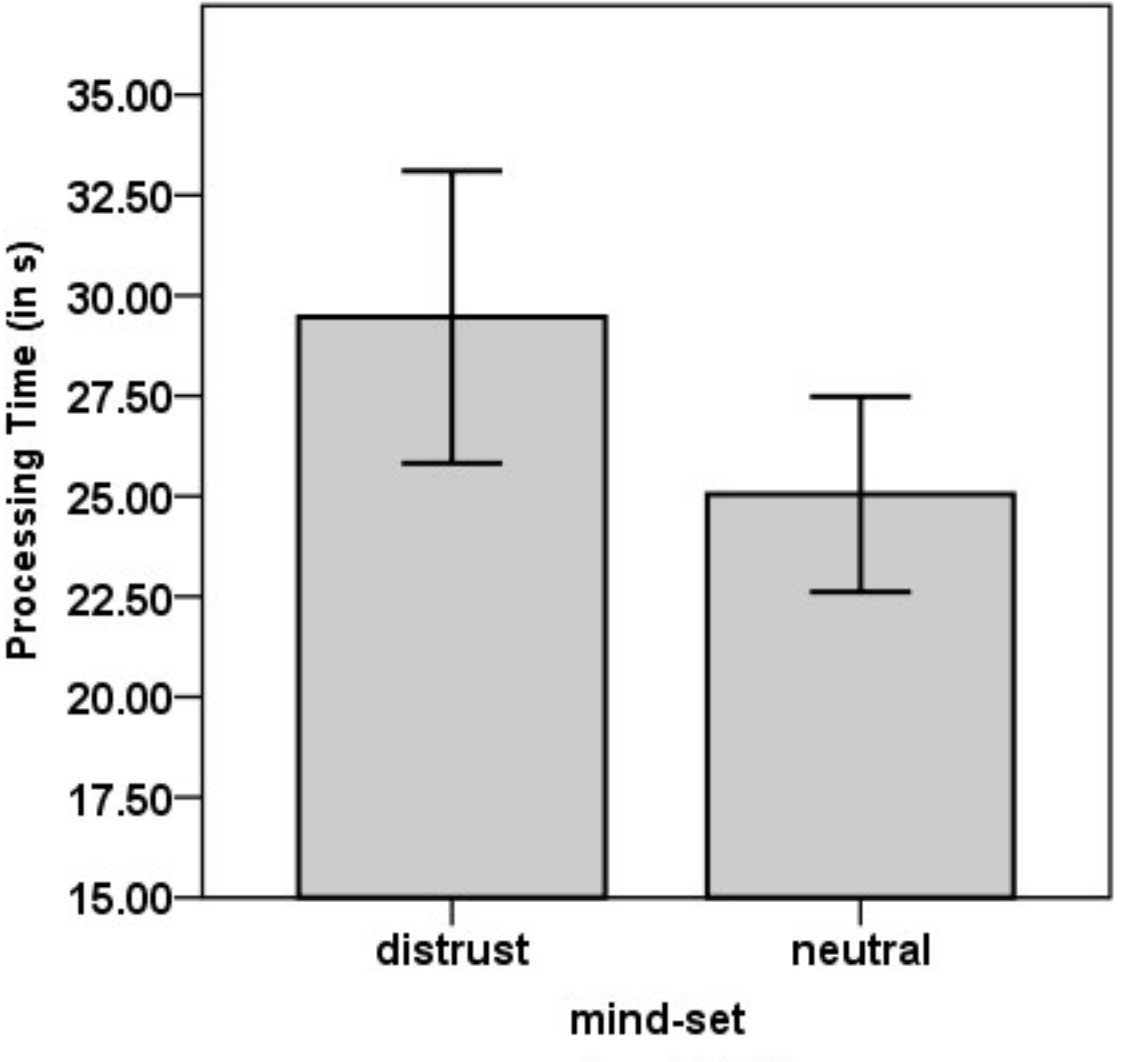

35.00
32.50
30.00
27.50
25.00
22.50
20.00
17.50
15.00
Processing Time (in s)
distrust
neutral
mind-set
error bars: 95% CI

Table 1

Content Analysis: Frequency of categorized responses in the different mind-sets.

Why is the report considered to be falsified or true?	
Response category	Frequency in %
neutral	
Contradiction/Consistency	16.70
Plausibility/Believability	40.00
Narrative style	30.00
Other	13.30
distrust	
Contradiction/Consistency	53.40
Plausibility/Believability	13.30
Narrative style	10.00
Other	23.30

Table 2

Average acceptance rates (M and SD) for the different syllogism types in the different mindsets

	mind-set		
Type of syllogism	distrust	neutral	overall
	M (*SD*) 95% CI [,]	*M (SD)* 95% CI [,]	*M (SD)* 95% CI [,]
valid – believable	.87 (.24) [.79, .94]	.79 (.26) [.70, .87]	.83 (.26) [.77, .88]
valid – unbelievable	.76 (.30) [.67, .85]	.71 (.34) [.60, .81]	.73 (.32) [.67, .80]
invalid – believable	.34 (.34) [.23, .45]	.53 (.35) [.42, 64]	.44 (.35) [.36, .51]
invalid - unbelievable	.15 (.21) [.09, .22]	.21 (.22) [.13, .27]	.18 (.22) [.13, .23]
overall	.53 (.13) [.49, .57]	.57 (.13) [.53, .61]	.55 (.13) [.52, .58]

Table 3

Average processing times (M and SD) for the different syllogism types for the different mind-sets

	mind-set		
Type of syllogism	distrust	neutral	overall
	M (*SD*)	*M (SD)*	*M (SD)*
valid – believable	284.20 (146.14)	242.59 (107.96)	263.40 (129.41)
valid – unbelievable	266.70 (115.40)	212.13 (99.74)	239.42 (110.66)
invalid – believable	330.39 (137.00)	286.17 (97.25)	308.28 (120.16)
invalid – unbelievable	297.30 (181.69)	261.03 (101.77)	279.15 (147.50)
overall	294.64 (116.88)	250.48 (77.98)	272.56 (101.22)