Distrust and the positive test heuristic:

Dispositional and situated social distrust improves performance on the Wason rule discovery task

Ruth Mayo[a], Dana Alfasi[a], & Norbert Schwarz[b]

a- The Hebrew University of Jerusalem
b- University of Michigan

Corresponding author: Ruth Mayo, Department of Psychology, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91905, Israel; email: ruti.mayo@huji.ac.il

Abstract. *Feelings of distrust alert people not to take information at face value, which may influence their reasoning strategy. Using the Wason (1960) rule identification task, we test whether chronic and temporary distrust increase the use of negative hypothesis testing strategies suited to falsify one's own initial hunch. In Study 1, participants who were low in dispositional trust were more likely to engage in negative hypothesis testing than participants high in dispositional trust. In Study 2, trust and distrust were induced through an alleged person-memory task. Paralleling the effects of chronic distrust, participants exposed to a single distrust-eliciting face were three times as likely to engage in negative hypothesis testing as participants exposed to a trust-eliciting face. In both studies, distrust increased negative hypothesis testing, which was associated with better performance on the Wason task. In contrast, participants' initial rule generation was not consistently affected by distrust. These findings provide first evidence that distrust can influence which reasoning strategy people adopt.*

Feelings of distrust alert us not to take information at face value and increase the likelihood that we will consider how things might differ from what meets the eye. This is reflected in increased accessibility of context-incongruent concepts and more unusual solutions on creativity tasks. Whereas exposure to a given concept (e.g., "temporary") usually facilitates the subsequent identification of related concepts that are congruent with the concept's meaning (e.g., "transitory;" for a review, see McNamara, 2005), the reverse holds under conditions of distrust. After incidental exposure to distrust-eliciting stimuli, people more quickly recognize "transitory" as a word when it follows the incongruent prime "permanent" than when it follows the congruent prime "temporary" (Schul, Mayo, & Burnstein, 2004). Subsequent work showed that induced distrust can facilitate access to multiple categories (Friesen & Sinclair, 2011; Mayer & Mussweiler, 2011), broaden category inclusiveness (Mayer & Mussweiler, 2011), increase the likelihood that people arrive at non-routine and creative solutions on problem-solving tasks (Mayer & Mussweiler, 2011; Schul, Mayo, & Burnstein, 2008) and foster a dissimilarity-focus (Posten & Mussweiler, 2013). Throughout, these findings are compatible with the assumption that distrust increases the accessibility of concept-

incongruent information, that is, information relevant to the basic question posed by the experience of distrust: How might things differ from what meets the eye?

Going beyond concept-accessibility effects, the present research explores whether distrust can also elicit shifts in people's reasoning strategies. We test this possibility in the context of a classic reasoning task that allows for the separation of hypothesis generation and hypothesis testing, namely, Wason's (1960) rule-discovery task. When asked to discover the rule underlying the number series 2-4-6, most people hypothesize a "+ 2" rule. When invited to test their hypothesis by generating a number series that will be marked as consistent or inconsistent with the correct rule, people overwhelmingly follow a positive testing strategy (Klayman & Ha, 1987) by generating a series that is consistent with their hypothesis (e.g., 8-10-12). The feedback they receive is positive and does not allow them to recognize their hypothesis is false – the correct rule is "any series of increasing numbers." Key to a positive testing strategy is a focus on instances that are congruent with one's hypothesis, namely, "instances in which the property or event is expected to occur (to see if it does occur)" or "instances in which it is known to have occurred (to see if the hypothesized conditions prevail)" (Klayman & Ha, 1987, p. 212). In the Wason rule-discovery task, a positive testing strategy results in hypothesis confirmation (Klayman & Ha, 1987; Oswald & Grosjean, 2004) and impairs discovery of the correct rule.

If distrust elicits a focus on potentially incongruent information, it may influence hypothesis generation, hypothesis testing, or both. At the hypothesis-generation stage, incidental distrust may prompt the generation of hypotheses that are incongruent with the apparent "+2" nature of the presented series. At the hypothesis-testing stage, distrust may prompt the generation of test series that are incongruent with the self-generated hypothesis, which would indicate a reduction in the prevalence of positive testing. Accordingly, the Wason rule-discovery task provides a unique opportunity to distinguish between these possibilities, unlike other creativity tasks, in which only the end result is available for inspection. Previous observations of increased non-routine solutions under distrust conditions (Mayer & Mussweiler, 2011; Schul et al., 2008) may either reflect that participants spontaneously

generated a non-routine solution, arrived at one by elaborating on possible alternatives for an initially generated routine idea, or both.

To date, studies have observed improvements in correct-rule discovery when participants were explicitly asked to discover two different rules (e.g., Gale & Ball, 2006; Gorman, Stafford, & Gorman, 1987; Tukey, 1986), when the series was presented as a counterexample to the experimenter's rule (e.g., Rossi, Caverni, Girotto, 2001), and when participants were explicitly given disconfirmation instructions (e.g., Gorman & Gorman, 1984). These conditions share that they explicitly invite participants to focus on how things may be different from the hypothesis they have in mind. We propose that the feeling of distrust is sufficient to induce the same shift. To test this prediction, study 1 relies on individual differences in distrust, whereas Study 2 primes distrust.

Study 1: Dispositional Distrust and Reasoning

Methods

Four hundred members of a diverse Israeli online panel (http://www.panel4all.co.il/panel), all with high school education or higher (mean age 28.8 years), completed allegedly unrelated tasks online. They first answered a Hebrew version of Yamagishi and Yamagishi's (1994) six item trust scale (e.g., "Most people are trustworthy"; 1=*strongly agree*, 7=*strongly disagree*).[1] Next, they were introduced to Wason's (1960) rule-discovery task, wrote down their hypothesis about the underlying rule, and generated six three-number series to test their hypothesis.[2] They then received feedback that indicated for each series whether it was congruent or incongruent with the underlying rule and subsequently generated a second hypothesis. A final questionnaire assessed participants' current mood; what they thought the purpose of the study was; whether they were familiar with the task;

---

[1] See online materials.

2.To ensure that the predictor (dispositional trust) is not affected by the dependent variable (Wason task) it was assessed first. This renders participants' own trust beliefs more accessible, which may increase their impact (Schwarz, 1987); however, it is unlikely to change participants' position along the trust dimension. Hence, the sequence provides a strong test of accessible differences in dispositional trust on reasoning.

whether they thought there was a trick to it; how confident they were about their first and second hypothesis; and whether they had completed the study alone (see online materials).

Results and Discussion

Forty-two participants did not report an initial hypothesis, nine had extensive missing data, and two did not work on the task alone; excluding these 53 participants leaves 347 participants for analysis. The six items of the trust scale were highly correlated (α = .862) and their mean score (Figure 1a) serves as the dispositional trust measure.

Rule generation. Trust was unrelated to rule generation. As is common for the Wason task, most participants generated a "+2" or "even numbers" rule. Only four participants thought of different rules, one each from each trust quartile.

Positive and negative testing. Participants' number series were coded as positive (0) or negative (1) tests of the participant's own hypothesis; for the "+2" hypothesis, the series 8-10-12 would be coded as a positive test, but 8-10-18 as a negative test.

 Overall, positive testing dominated and 155 of 347 participants did not include a single negative test in their number series. This highly skewed distribution (Figure 1b) calls for a dichotomized analysis (MacCallum, Zhang, Preacher & Rucker, 2002) that compares participants who generated at least one negative test with those who generated none. A logistic regression indicates that trust predicts the choice of testing strategy; χ^2 = 5.47, *df* = 1, *p* <0.05, Phi=0.13, b=0.27. Whereas more than two thirds (68.82%) of the participants in the lowest trust quartile generated *at least one* negative test, less than half (48.86%) of those in the high trust quartile did so; χ^2 = 7.44, *df* = 1, *p*<.007, Phi=0.20 (Figure 2). This supports the prediction that low-trust participants are more likely to generate negative tests than high-trust participants. This relationship is also reflected in the observation that those who generated at least one negative test were significantly lower in trust (M=3.98) than those who did not (*M* = 4.23), *t*(345)=2.37, *p*< .05, *d*=0.25.

Correct solutions. Overall, negative testing increased the likelihood of arriving at the correct rule when participants reported their second hypothesis. Of the 192 participants who generated at least one negative test, 33 (17.19%) arrived at the correct rule, whereas only two (1.29%) out of the 155 participants who generated only positive tests did so; χ^2 = 23.89, *df* = 1, *p*

< .0001, Phi=0.26. Nevertheless, the observed higher prevalence of negative testing under low trust did not translate into significantly better performance at this step; χ^2 = 0.69, df = 1, p = .405, for the lowest trust quartile vs. the high trust quartile participants.

   Additional analyses. Participants who reported a negative test did not differ in mood (M = 6.69) from those who did not (M =6.72); t(346)=0.03, p=0.85. Negative testers reported less confidence in their initial and final hypothesis (p's < .005) and thought it more likely that the series involved a "trick" (t(346)=4.32, p<0.0005). However, the four quartiles of trust did not differ in whether they thought the task involved a "trick", p=0.18.[3] Note, however, that these variables were assessed after participants reported their final hypothesis and may have been affected by the feedback they received on their number series.

Study 2: Situated Distrust and Reasoning

   Whereas Study 1 relied on individual differences in trust, Study 2 manipulated trust by exposing participants to a single trustworthy or untrustworthy face as part of an unrelated task. Previous research documented the effectiveness of this manipulation (Schul et al. 2004, 2008).

Methods

   Forty undergraduates participated for course credit and were randomly assigned to a trust or distrust condition. They read: *"This study has multiple parts. First, we would like you to look at a face of a person and form an impression of him. Later, we will ask you to remember this person and the impression you formed about him. To make the memory task harder, we will ask you to work on an unrelated task in between so that a little time can pass."* The next page presented a male face (Figure 3) that served as a trust or distrust cue (taken from Schul et al., 2004). The alleged filler task was Wason's (1960) rule-discovery task. Participants received the number series 2-4-6, wrote down their hypothesis about the underlying rule, and generated six three-number series to test their own hypothesis (see online materials for instructions).

   Completing the person-memory cover story, participants then listed the thoughts that had come to mind when they saw the face, and reported their impression by checking descriptive adjectives on a list (smart, warm, deceptive, happy, shy, trustworthy, independent,

---

[3] See online materials for more detail.

romantic, competitive, content, sociable, young); the terms "deceptive" and "trustworthy" served as a manipulation check. Finally, the experimenter marked each number series as correct or incorrect and asked participants to write down what they now believed to be the correct rule.[4]

Results and Discussion

Manipulation check. Confirming the effectiveness of the trust manipulation, 70.0% (14 of 20) of the participants in the distrust condition checked the term "deceptive" for the narrow-eyed face; no participant checked "trustworthy." In the trust condition, 25% (5 of 20) of the participants checked "trustworthy" and only 10% (2 of 20) checked "deceptive." We excluded the latter two participants because their responses to the face indicated distrust; however, their inclusion would not change the conclusions.

The finding that more participants endorsed "deceptive" in the distrust condition than "trustworthy" in the trust condition presumably reflects that "deceptive" is a marked characteristic that is worth reporting, whereas moderate trust is the default in most human interactions and hence less noteworthy.

Rule generation. All participants in the distrust condition listed the "+2" or "even numbers" rule as the first rule that came to mind, whereas four participants (22.22%) in the trust condition generated a different rule ($\chi^2 = 5.71$, $df = 1$, $p < .05$). The responses in the trust condition are consistent with the familiar observation that about 80% initially report a "+2" hypothesis (Tukey, 1986; Wason, 1960; Wharton, Cheng, & Wickens, 1993); distrust further increased the dominance of this rule. Thus, neither dispositional (Study 1) nor induced distrust elicited more divergent rule generation.

Positive and negative testing. Replicating Study 1, distrust increased the likelihood of negative testing. As shown in Figure 4, only 3 of 18 (16.67%) participants exposed to a trust-inducing face generated *any* number series suitable as a negative test, whereas 12 of 20 (60%) participants exposed to a distrust-inducing face did so ($\chi^2 = 7.44$, $df = 1$, $p<.007$, Phi=0.44).[5]

---

[4] Following Gorman and Gorman (1984), the procedure of having participants write down all series and receive feedback at the end was chosen to avoid conveying disconfirmatory information during task performance.

5. As in Study 1, the distribution of the negative test measure was extremely highly skewed

Thus, incidental distrust tripled the proportion of participants who generated at least one negative test. Overall, 27.5% of all series in the distrust condition, but only 7.4% in the trust condition, were coded as negative tests of participants' own hypothesis; $t(36)= 2.33$, $p < 0.05$, d=0.75.

   Correct solutions. As in Study 1, negative testing increased the likelihood that participants' second hypothesis offered the correct solution. Given the close relationship between induced distrust and negative testing, 30.0% of the participants in the distrust condition arrived at the correct rule, whereas only 5.56% of the participants in the trust condition did so; $\chi^2 = 3.76$, $df = 1$, $p = .052$.

General Discussion

   Our findings converge on the conclusion that chronic (Study 1) or temporary (Study 2) distrust increases the use of negative hypothesis testing strategies on reasoning tasks. A single exposure to a distrust-eliciting face was sufficient to triple the rate of participants who used a negative testing strategy, from less than 20% under trust to 60% percent under distrust conditions. Negative testing, in turn, was associated with improved performance on the Wason (1960) rule identification task in both studies.

   In contrast, we found little if any influence of distrust at the hypothesis-generation stage. Most importantly, the distrust-elicited increase in the accessibility of incongruent information observed on lexical decision tasks (Schul et al., 2004) was not paralleled by the generation of initial hypotheses that were incongruent with "what meets the eye." If anything, participants in a distrust context (Study 2) were more rather than less likely to focus on the salient regularity of the "+2" pattern in the series presented to them, but then proceeded to test that hypothesis with a negative testing strategy.

   In combination, these findings highlight the importance of identifying the stage at which distrust exerts influence. Although distrust can increase the accessibility of incongruent information as assessed by lexical decision tasks (Schul et al., 2004), the initial idea-generation stage is not the only route to the higher creativity observed by Mayer and Mussweiler (2011), or the improved problem solving observed by Schul et al. (2008). Instead, distrust can also

improve performance by facilitating incongruent elaborations of the initially generated ideas at the evaluation and testing stages. Both effects of distrust are adaptive responses to the key problem signaled by distrust: things may be different than they seem.

Cowley and Byrne (2005) suggested people might be more inclined to falsify others' hypotheses than their own. The authors compared a condition in which an imaginary other proposed the "+2" rule with a condition in which participants were asked to think of the "+2" rule as a hypothesis they had offered themselves (though the experimenter provided the rule in all conditions). Under these conditions, participants were non-significantly more likely to use negative tests when the experimenter attributed the rule to someone else rather than to them, leading the authors to suggest that "participants may be able to rely on a falsification strategy in a rational way to test somebody else's hypothesis but not their own" (Cowley & Byrne, 2005, p. 517). Similarly, Mercier and Sperber (2011) proposed that reasoning serves argumentative functions and "does exactly what can be expected of an argumentative device: Look for arguments that support a given conclusion, and, ceteris paribus, favor conclusions for which arguments can be found" (p. 57). From this perspective, negative testing should be particularly unlikely when people evaluate self-generated hypotheses. Nevertheless, a majority of the participants who were either chronically low in trust (Study 1) or exposed to a distrust-eliciting face (Study 2) attempted to falsify their own hypotheses in the present studies. One may argue that this reflects suspicions about the task presented to them, yet level of trust was unrelated to participants' impression of whether the task entails a trick (Study 1), suggesting that the cognitive strategies engaged by distrust extend beyond the examination of information explicitly provided by others.

This conjecture is consistent with the assumption that thinking is tuned to meet one's needs and that feelings play a key role in this process (for reviews, see Schwarz, 2002, 2012; Smith & Semin, 2004). Distrust signals a situation in which one may be ill advised to take information at face value. Under such conditions it is wise to entertain how things may differ from what meets the eye and to carefully examine one's own reasoning. Whether such distrust-induced shifts in reasoning are beneficial should depend on the task. When the first thought that comes to mind is likely to be misleading, as is the case for Wason's (1960) rule-discovery

task, distrust is likely to help. On other tasks, distrust may induce its own biases, leading perceivers to erroneously "detect" meanings that were not intended, as illustrated by the adverse effects of distrust in negotiations (for a review, see Thompson, 2005). Exploring the influence of distrust across a variety of tasks, including the diverse judgment and decision phenomena often subsumed under the label "confirmation bias" (Gilovich, 1991; Nickerson, 1998), provides a promising avenue for future research at the interface of situated cognition, individual differences, emotion, and reasoning; such extensions may be complemented by exploring the influence of other feelings, which provide different information and hence invite different processing strategies (Blanchette & Richards, 2010; Han, Lerner, & Keltner, 2007; Schwarz, 2012; Tiedens & Linton, 2001).

References

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgment, decision making and reasoning. *Cognition & Emotion, 24*, 561-595

Cowley, M., & Byrne, R. M. J. (2005). When falsification is the only path to truth. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), Proceedings of the 27th Annual Conference of the Cognitive Science Society (pp. 250-255). Stresa, Italy: Mahwah, NJ: Erlbaum.

Friesen, J., & Sinclair, L. (2011). Distrust and simultaneous activation of multiple categories. *Social Psychological and Personality Science, 2*, 112-118.

Gale, M., & Ball, L. J. (2006). Dual-goal facilitation in Wason's 2-4-6 task: What mediates successful rule discovery? *Quarterly Journal of Experimental Psychology, 59*, 873–885.

Gilovich, T. (1991). *How we know what isn't so.* New York: Free Press.

Gorman, M., & Gorman, M. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology, 36A*, 629-648.

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual

hypotheses on a more difficult version of Wason's 2-4-6 task. *Quarterly Journal of ExperimentalPsychology, 39A*, 1–28.

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology, 17,* 158-168.

Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review, 94*, 211–228.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19-40.

Mayer, J., & Mussweiler, T. (2011). Suspicious Spirits, Flexible Minds: When Distrust Enhances Creativity. *Journal of Personality and Social Psychology*, *101*, 1262-1277.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34,* 57-111

Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220

Oswald, M.E., & Grosjean, S. (2004). Confirmation bias. In Pohl, R. F. (ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (pp. 79-96). New York: Psychology Press.

Posten, A.-C., & Mussweiler, T. (2013, June 17). When Distrust frees your mind: The stereotype-reducing effects of distrust. *Journal of Personality and Social Psychology*, doi 10.1037/a0033170

Rossi, S., Caverni, J. P., & Girotto, V. (2001). Hypothesis testing in a rule discovery problem: When a focused procedure is effective. *Quarterly Journal of Experimental Psychology, 54A,* 263-267.

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.*

Schul, Y., Mayo, R., & Burnstein, E. (2008). The value of distrust. *Journal of Experimental Social Psychology, 44,* 1293-1302.

Schwarz, N. (1987). Geschlechtsrollenorientierung und die Einstellung zu Gewalt gegen Frauen: Informationsaktivierung als Alternative zu ex post facto - Versuchsplänen. [Gender role orientation and attitudes toward violence against women: Priming procedures as an alternative to ex post facto - designs.] *Psychologische Rundschau, 38*, 145 - 154.

Schwarz, N. (2002). Situated cognition and the wisdom of feelings: Cognitive tuning. In L. Feldman Barrett & P. Salovey (Eds.), *The wisdom in feelings* (pp.144-166). New York: Guilford.

Schwarz, N. (2012). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & E. T. Higgins (eds.), *Handbook of theories of social psychology* (pp. 289-308). Los Angeles, CA: Sage.

Smith, E.R., & Semin, G.R. (2004). Socially situated cognition: Cognition in its social context. *Advances in Experimental Social Psychology, 36, 53-117.*

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, *81*, 1-16.

Thompson, L.L. (2005). *The mind and heart of the negotiator.* New York: Pearson.

Tukey, D. D. (*1986). A philosophical* and empirical analysis of subjects' modes of inquiry in Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology, 38A*, 5–33.

Wason, P. C. (1960). On the failure to eliminate hypothesis in a conceptual task. *Quarterly Journal of Experimental Psychology, 14, 129-140.*

Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis-testing strategies: Why two goals are better than one. *The Quarterly Journal of Experimental Psychology, 46A*, 743-758.

Yamagishi, T. & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129-166.

Figure 1 (a): Mean trust measure distribution (Study 1)
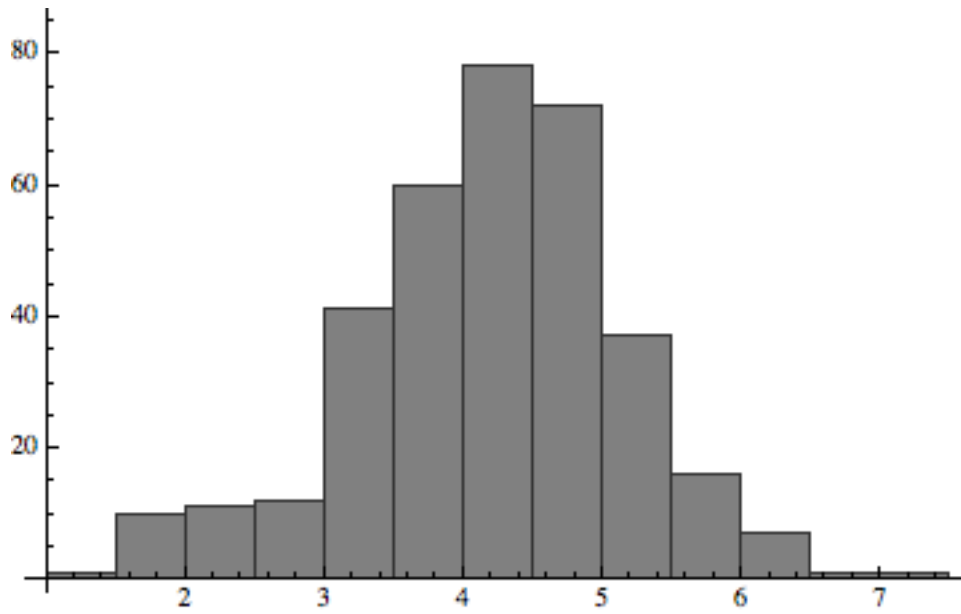
- *Y axis = number of participants, X axis =mean of trust measure.*

Figure 1 (b): Mean negative testing measure distribution (Study 1)
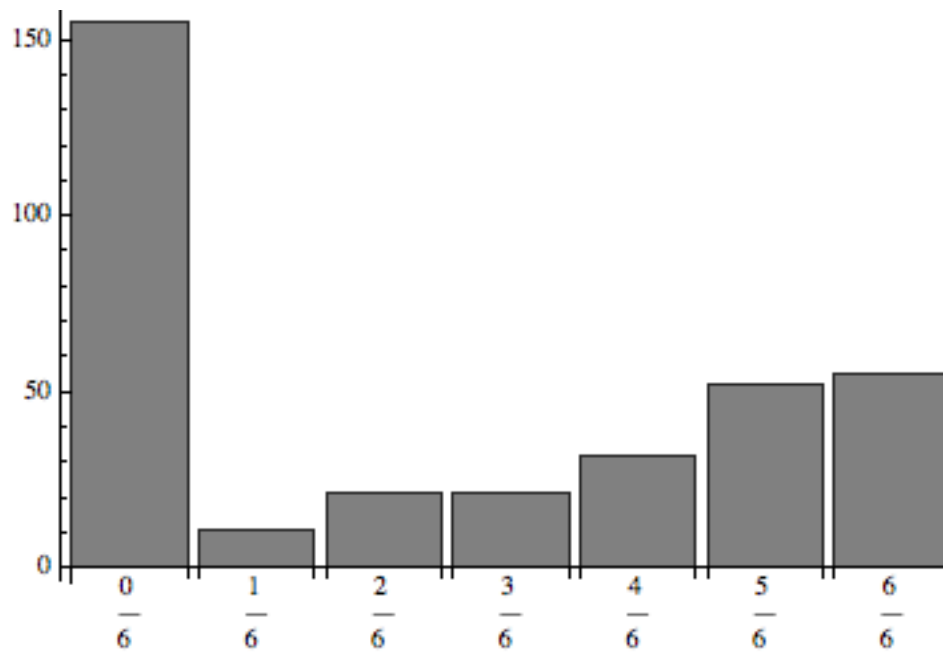
- *Y axis = number of participants, X axis = number of series out of six that were negative testing.*

Figure 2. Percentage of participants who generated at least one negative test as a function of dispositional trust (Study 1)
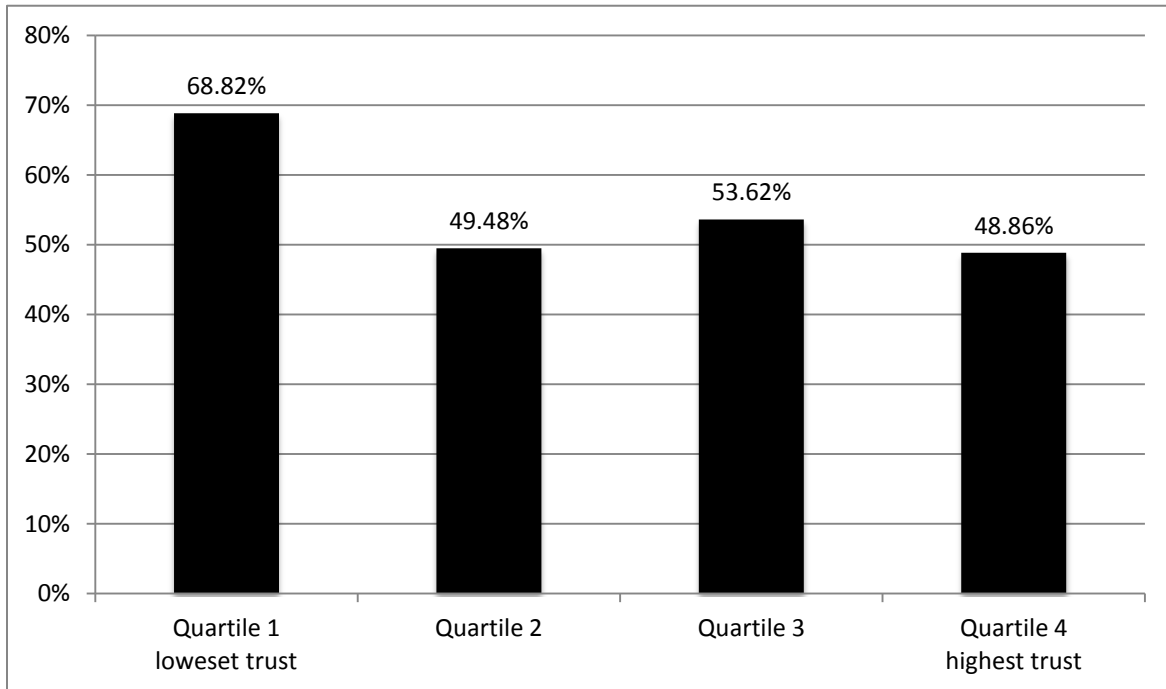
Figure 3. Trust- and distrust-eliciting face primes

Trustworthy face

Untrustworthy face

Figure 4: Percentage of participants who generated at least one negative testing number series as a function of trust versus distrust context (Study 2)