



Högskolan  
Kristianstad

Högskolan Kristianstad  
291 88 Kristianstad  
044 250 30 00  
[www.hkr.se](http://www.hkr.se)

Independent project (degree project), 15 ECTS,  
Master's degree in Psychology  
Spring Semester 2022  
Faculty of Education

# Further Insights into Truth-Default Theory

## A Deception Detection Experiment

Daniel M. O. Malmgren

**Author**

Daniel M. O. Malmgren

**Title**

Further Insights into Truth-Default Theory: A Deception Detection Experiment

**Supervisor**

Marcus Johansson

**Examiner**

Peter Jönsson

**Abstract**

People seem to rarely consider whether a communicated narrative is deceptive. The aim of this study was to replicate previous findings about the truth-default theory, and to provide further insights into the effects of the truth-default state in processing another's communication. The Truth-Default theory provides a framework for understanding people's tendency to operate on a presumption of honesty. The experiment reported in this study used video-recorded narratives and four between-subject factors in investigating the effects of different types of triggers on the truth-default state. A total of six hypotheses were studied: hypothesis one through three were about replicating previous findings on the occurrence and frequency of thoughts pertaining to veracity or deception, the levels of truth-bias, deception detection accuracy and the importance of plausibility in comparison to actual veracity. Hypothesis four through six predicted effects of semantic priming, implausible message content and possible mediating effects on ratings of credibility and reliance through liar representativity and plausibility. In summary, prior findings on the truth-default state replicated and were corroborated. The outcome of the experiment done by 106 participants online indicates that people tend to passively assume that others are honest and that communicated narratives are reliable. The result calls for renewed attention to the risks of malign information influence posed by junk news producers and others. It also highlights the importance of educating in and developing source criticism skills.

**Keywords**

Lying, deception, bias, tendency, truth-default theory, source criticism

### Further Insights into Truth-Default Theory: A Deception Detection Experiment

In 2017, a college student in the United States decided to join a demonstration holding a placard with the text “Birds aren’t real”. This conspiracy theory states that all birds are dead, and that the birds we see flying are drones controlled by the American government (Lorenz, 2021). Some people believe in this satirical conspiracy theory, despite the blatant violation of common sense and a lack of adequate evidence making it a clear violation of Grice’s maxim on quality (Grice, 1989). Truth-Default theory (Levine, 2014; Levine, 2020) is a theory on deception that provides a framework to understand why people presume that a communicated narrative is true.

The present study focuses on deception and people’s tendency to believe what’s being communicated to them. Deception occurs when someone “intentionally, knowingly, or purposefully” (Levine, 2014, p. 380) misleads another person. Lying is a type of deception that occurs when a sender communicates something which the sender believes or knows to be untrue with the purpose of deceiving the receiver (Bok, 1979; Dynel, 2011). Serota et al. (2010) showed that the prevalence of lying is not evenly distributed in a population. A small number of people communicates most of the deception efforts. Lies are on average told around two times a day and without much effort (Serota et al., 2010; DePaulo et al., 1996). Communicators of lies are usually not worried about getting caught (DePaulo et al., 1996).

In addition to the finding that people seldom lie, empirical evidence show that people rarely assess the veracity of what’s being communicated to them (Clare & Levine, 2019; Levine et al., 2021). This occurs for an example even though internet users are concerned about false information and fraud (The Lloyd’s Register Foundation World Risk Poll Report, 2019). In connection with this, Levine et al. (2021) used a novel angle to investigate the occurrence of spontaneous veracity assessments. The participants of that study ( $N = 163$ ) viewed one out of six videos on YouTube. The different videos varied in credibility (actual factual merit), and credulity (surface appearance of credibility). Except for a satirical news story, not more than 15.7% of the participants mentioned any thoughts on deception or veracity.

The automatic tendency to believe communication as honest is called truth-bias (McCornack & Parks, 1986). The use of the term bias can be somewhat misleading given that biases are often viewed as maladaptive (Levine, 2020). The tendency to believe others can however be modelled on how implicit biases operate, namely without intention and conscious awareness (Bargh, 1989). Biased judgements based on heuristics can be an adequate response in a given context (Dutilh Novaes & Veluwenkamp, 2017). Tversky and Kahneman (1974) noted in their article that “heuristics are quite useful, but sometimes they lead to severe and systematic errors” (p. 1124). Truth-Default theory (TDT; Levine, 2014; Levine, 2020) is a theory on deception that accounts for truth-bias and regards it as a key concept. TDT holds that efficient communication is facilitated through a passive presumption of honesty, a truth-default state of mind. The truth-default state is most of the time harmless, i.e., not maladaptive given that most people are honest most of the time (Levine, 2010; Halevy et al., 2014; Daiku et al., 2021). Truth-bias is empirically quantified as the percentage of messages that are judged as true in any given setting (McCornack & Parks, 1986).

The existence of a truth-default state is as noted by its author (Levine, 2014) supported by other theories, such as the pragmatic theory on conversation developed by Grice (1989). People act cooperatively in conversations to be understood. One of the Gricean maxims (read assumptions) is on quality, which entails sincerity and truthfulness. Conversational partners rely on the same assumptions, which the sender can use to generate implicatures, i.e., implying without saying. The receiver infers a sensible message based on above-mentioned assumptions. This allows the conversation to flow without explicitly stating everything (Grice, 1989; Dynel, 2011; Fallis, 2012).

A second theoretical basis of a truth-default state is the Spinozan theory. Receivers of any type of communication entertains propositions, i.e., the content of the communication. The Spinozan theory posits that an entertained proposition is a believed proposition (Gilbert et al., 1990). People cannot withhold a judgement on the veracity of the communicated content. A message receiver will first believe the message to be true, and then consider if it's deceptive (Mandelbaum, 2014). People seeing a placard with the text "Birds aren't real" must believe it first before being able to discard it as incorrect or deceptive.

The truth-default state explains the tendency to believe what's being communicated. To use the now famous vocabulary from the dual-process theory promoted by Tversky and Kahneman (1974), people need effortful and controlled thinking from system 2 to actively circumvent the truth-default in system 1. Stanovich and West (2000) introduced system 1 and system 2, referring to Gricean theory in describing system 1 as automatic, undemanding and one used for communication. System 1 processes use conversational implicatures, even in situations lacking the typical characteristics of a conversation. System 2 is demanding and used for analytical and critical thinking. Where system 1 is personalized, system 2 is depersonalized and decontextualizes the judgements under consideration. The TDT view on triggers for veracity assessments can be phrased in analogy with the system 1 and system 2 approach. At least one sufficiently potent trigger is needed to go from the truth-default state in system 1 to suspicion and the critical thinking of system 2. Moreover, a second trigger is needed to move from suspicion to active disbelief (Levine, 2020). Examples of triggers includes but are not limited to a known motive for deception, incoherent message content, known facts that contradicts the message content, a warning from a third party of possible deception, or dishonest demeanor (Levine, 2020). The truth-default state is adaptive, and people can abandon this state given sufficiently potent triggers (Levine, 2014).

TDT consists of different modules and one module is on sender honest demeanor. Demeanor is "a constellation of behaviors that conveys to others a sense of the sender's character" (Levine et al., 2011, p. 378; Goffman, 1956). While Grice (1989) refers to violations of implicatures, Goffman (1956) refers to infractions of rules of conduct. An infraction has a negative impact on the sense of the person's character (Goffman, 1956). In a series of five experiments, Levine et al. (2011) varied sender demeanor to create demeanor-veracity matched and demeanor-veracity mismatched conditions. Sender demeanor explained approximately 98% of the variance in deception detection accuracy. People trust and rely more on some, than in others dependent on that person's demeanor (Levine et al., 2011; Bond & DePaulo, 2008; Hartwig & Bond, 2011). Tversky and Kahneman labelled this route to biased judgements the representativeness heuristic (Tversky & Kahneman, 1974).

Reliable ways of detecting deception have been researched within different contexts, mirroring different ecologies of the task. Meta-analytical work shows an average accuracy of 54% in correctly assessing veracity (Bond & DePaulo, 2006). Accuracy refers to the percentage of participants that correctly assess a true narrative as truthful, and a deceptive narrative as a lie. Psychophysiological and neural measures have been promoted in the forensic context. However, the validity of those measures relies on the validity of a question format that isolates deception (Meijer & Verschuere, 2018). Meta-analysis has furthermore showed minimal effects of age, education level, work experience and sex on deception detection performance (Aamodt & Custer, 2006). Granhag et al. (2007) have developed the Strategic Use of Evidence approach, which minimizes the reliance on cues to deception and instead uses information from third parties to detect deception. This approach has proven successful in the forensic context (Hartwig et al., 2006; Luke et al., 2016). Researchers have also distilled objective cues for plausibility with the purpose of assessing veracity. Vrij et al. (2020) re-analyzed five datasets. They found that frequency of details, complications, and verifiable sources were positively correlated with plausibility.

The phenomenon of truth-bias has been acknowledged in studies and in meta-analytical work on deception detection (Zuckerman, et al., 1981; Levine et al., 1999; DePaulo & Bond, 2008; Vrij, 2000;

Swol et al., 2015). The human tendency to believe can have an undue impact on deception detection experiments. An instruction, prompt, or a blatant trigger to assess veracity can be conceptualized as a threat to the validity of a deception detection experiment which has different implications on different ecologies (Levine, 2018). The forensic context is very different from everyday conversation and the current study aims to mirror everyday communication, not deception detection in the forensic setting.

The threat of an instruction to assess veracity can be phrased as a question: Would the participant consider deception if not prompted by the researcher to make that kind of judgement? Clare and Levine (2019) investigated whether prompt versus no prompt affected the occurrence and frequency of veracity assessments. In two mixed design experiments, each participant evaluated four types of messages, one plausible truth, one implausible truth, one plausible lie, and one implausible lie. Participants in the prompted condition were instructed to make a dichotomous choice between truth or lie. Participants in the unprompted condition were asked to type in any thoughts about the confederate's answers. The prompt or no prompt was used as an independent group factor. Deception detection prompting, message content veracity and plausibility were repeated factors. Example of a plausible story included favorite food, and an implausible story was about delivering a baby in a traffic jam. Participants in the first experiment saw video-recorded interviews. Participants rarely assessed veracity unless prompted to do so. The second experiment replicated the results using an interactive setting. Instead of watching videotapes, participants did an interview with a confederate in face-to-face dyads. Message veracity and plausibility was again crossed. Participants in the unprompted condition were given the same thought-listing question as in experiment one, this time audio-recording their answers. Veracity assessments were few.

Both experiments reported in Clare and Levine (2019) employed a 50/50% base-rate of truths and lies. Experiment one was done in a classroom setting and experiment two in a laboratory with small rooms. Less than 5% of the participants in the unprompted condition in experiment one mentioned thoughts about veracity, compared to almost 40% of the participants in the prompted condition. Even in the presence of a prompt, the overall truth-bias in experiment one was 58.42%. For experiment two, only .57% in the unprompted and 8.46% in the prompted condition explicitly mentioned thoughts about deception or veracity. The truth-bias was even higher in experiment two resulting in 88.43% of the participants believing in the messages. The participants were more likely to mention veracity about implausible than plausible content, making plausibility a bigger contributor to mentions of veracity than actual veracity.

Clare and Levine (2019) concluded that "absent prompting, priming, or another trigger, the possibility that a message might be deceptive simply does not come to mind" (p. 302). The present study is a partial replication of the study by Clare and Levine (2019). The primary aim is to replicate the findings from Clare and Levine (2019) and in part Levine et al. (2021).

The current study also investigates whether semantic priming has the potency of a trigger for participants to make veracity assessments. Kahneman (2003) proposed that the accessibility of an intuitive thought is decided by the features of the cognitive mechanisms and the characteristics of the event and stimuli that evoke it. A primed stimuli can increase cognitive accessibility, i.e., making the primed trait terms more salient and more easily accessible. This in turn can affect thought and behavior in a prime-congruent manner (Higgins & Eitam, 2014; see Bargh, 2006 for an overview). Previous experiments on semantic priming have shown effects on participants judgements and behaviors (Mussweiler & Strack, 1999; Was et al., 2019; Weingarten et al., 2016). Higgins (1977) primed participants with either positive or negative trait terms. The prime affected the participants impression of the targeted person. Bargh et al., (1996) primed participants with an elderly stereotype making them walk slower in comparison to a control group. The Bargh et al. (1996) experiment and others on priming have failed to replicate (Doyen et al., 2012; Harris et al., 2013; Nelson et al., 2013).

The experiment reported in this study used the pretense of a memorization task to semantically prime the participants with terms related to deceit or truthfulness, which aimed to increase the saliency and accessibility of such trait terms. Hence, priming was tried as trigger one in the first phase of the experiment. Implausible content was used as trigger two in phase two of the experiment. Levine (2020) notes that people possess a kind of general knowledge about what is usual and or typical in different contexts, allowing inference about what is plausible and implausible. Sorensen (2007) used the concepts of narrow and wide plausibility in distinguishing between different types of lies. Sorensen argues that a sender using wide plausibility is stating something that lacks credibility relative to what one can offer as evidence.

The current study seeks to broaden the demographical basis on which inferences about the truth-default state can be made (Clare & Levine, 2019; Levine et al., 2021). The experiment used in this study had four phases. The participants were in the first phase shown 30 words that in part provided the semantic priming. The participants were then in the second phase randomized to see one of four possible message conditions, a plausible truth, an implausible truth, a plausible lie, or an implausible lie. The participants were in the third phase instructed to type in thoughts pertaining to the message stimulus and to rate the content of the video clip on different dimensions of veracity using Likert-scale questions. The participants were then in the fourth and last phase shown a new list of words and asked to select which of them had been on the phase one list.

The video clips in the experiment reported here were made to remove confounding effects of diversity in content irrelevant for the actual veracity assessment. The YouTube clips used in Levine et al. (2021) were as pointed out by the authors of that study very different from one another, confounding comparisons between the different video clips. The video clips used in this study were furthermore made to mimic spontaneous, rather than planned narratives. The confederates were given a limited amount of time to prepare. All clips were filmed in one shot.

The questions relating to dimensions of veracity were designed not to trigger any thoughts on veracity that might not have occurred otherwise. The measured dimensions were the confederate's credibility, the reliability of the statement, liar representativity, and message content plausibility. Credibility was represented as a question about trust. High levels of trust within relationships can have a detrimental effect on deception detection accuracy (McCornack & Parks, 1986). Distrust and suspicion decrease truth-bias but increase the number of erroneous veracity assessments of truths (Levine et al., 1999). The question on statement reliability provides an indirect measure to see if the narrative was believed. The question on liar representativity takes the representativeness heuristic into account. The participants rated how characteristic the behavior of the confederate was of a liar. This question controls for the assumption that deceptive demeanor can trigger veracity assessments. Given that plausibility contributed to veracity assessments in Clare and Levine (2019) the participants in this study were also asked to rate the plausibility of the message content. The question was phrased in line with a definition by Leal et al. (2019): "How likely is it that the activities happened in the way described?". Hence, plausibility was used both as a factor and as a dependent variable.

The primary research question is in line with the research done by Clare and Levine (2019); how robust is the truth-default state in processing another's communication? This question guided the formulation of the six hypotheses for this study.

**H1** - The first hypothesis predicts that between 4% and 20% of the participants types in deception or veracity related thoughts. Clare and Levine (2019) reported that 4.23% of the participants made a spontaneous veracity assessment after having seen a video-recorded narrative. Levine et al. (2021) reported that fewer than 20% of the participants had expressed skepticism or suspicion. Apart from a satirical news video which made 27.5% of the participants correctly labelling it as satire or parody.

**H2** - The second hypothesis predicts that the level of truth-bias will be the lowest for implausible lies and the highest for plausible truths, while accuracy will be lowest for plausible lies and highest for plausible truths. Whereas, the first hypothesis relates to the frequency and occurrence of thoughts pertaining to veracity assessments, the second hypothesis refers to the levels of accuracy and truth-bias. Clare and Levine (2019) reported that an overall of 58.42% of the participants believed in the narrated contents. Implausible lies had the lowest level, with 37.50% of the participants believing in the narrated content. Plausible truths had the highest percentage of participants believing in the content, reaching 72.92%. Participants viewing plausible lies had the lowest accuracy, 38.89% of the participants correctly assessed the content as false. Participants viewing the plausible truthful content had the highest accuracy, 72.92% of the participants made a correct assessment. Levine et al. (1999) showed in three experiments that accuracy rates for truths were greater than chance, while accuracy for lies were below chance. Levine et al. labelled this phenomenon of higher accuracy rates for truths than lies the “veracity effect”, which is also the name of a TDT module.

**H3** - The third hypothesis is that implausible content independent of actual veracity of the content generates the highest frequency of deception or veracity relevant thoughts in the thought listing phase. These three hypotheses together test the robustness of the results obtained in the previous studies (Clare & Levine, 2019; Levine et al., 2021) and whether the results replicate when using a demographically different sample.

**H4** - The fourth hypothesis is that semantic priming of words about deceit together with implausible content increases the occurrence and frequency of veracity assessments in comparison to no priming and the priming of words about truthfulness. The priming of terms related to deceit aimed to provide the first trigger for participants to become suspicious. The no priming condition acted as a control group, also enabling a direct comparison to previous findings. The priming of words that increases the saliency of terms regarding truthfulness allows a directional comparison of priming effects. Implausible content was used as a second trigger and purported to move the participant from suspicion to active disbelief.

**H5** - The fifth hypothesis is that message content plausibility takes precedence over message content veracity in differentiating the participants ratings of the communicator’s credibility, statement reliability, representativeness, and content plausibility. This hypothesis aimed to assess how actual veracity and message content plausibility jointly differentiates participants ratings on the different dimensions of veracity. The plausibility ratings were treated separately in assessing the above-mentioned influence. This was done given the impact of plausibility on the results reported by Clare and Levine (2019).

**H6** - The sixth and final hypothesis is that ratings of credibility and ratings of statement reliability is mediated by ratings of representativeness as a liar and of message content plausibility. Given that some people display a more credible demeanor than others, demeanor-based judgements of veracity are fallible. Sender demeanor is nevertheless still one of the main determinants of deception judgements (Bond & DePaulo, 2008; Bond et al., 2015). Also, plausibility made a discernable impact on veracity assessments in Clare and Levine (2019). This gives a reason to predict that plausibility affects the reliability of a given statement. This sixth and last hypothesis tests the relationship between a person’s credibility and the reliability of a statement coming from that person, and whether that relationship can be explained by representativeness as a liar or message plausibility.

## **Method**

### **Participants**

A chain-referral sampling method was employed to reach out to different social networks. The participants were mainly recruited through email and consisted of a convenience sample. No

compensation was given for participation. 111 Swedish participants finished the experiment. A technical difficulty resulted in a skewed randomization and a base-rate of truthful content at 58.5% and deceptive content at 41.5%. This also skewed the base-rate of plausible (41.5%) and implausible content (58.5%). A small number ( $n = 4$ ) of participants that viewed one of the plausible lie videos were excluded because of the skewed randomization. One participant knew the confederate which resulted in exclusion. In the end, data from 106 participants were analyzed ( $M_{age} = 46$ ,  $SD = 13.9$ , range = 23-80 years, 48% women). The technical difficulty led to 30 participants being randomized into the control condition, 38 to the truthfulness prime condition and 38 to the deceit prime condition. The participants educational background ranged from elementary school to PhD, and 69.8% held a bachelor's or higher degree. Most of the participants were employed (78%), the remaining were either students, retirees or unemployed. None of the background variables made a discernable difference on the answers. The background variables will not be treated any further.

### **Design**

The experiment consisted of four between-subjects factors. The first factor was semantic priming consisting of three levels. Level one was the control condition with no priming, level two primed participants with terms regarding truthfulness, and level three primed participants with terms regarding deceit. The second and third factor both had two levels and regarded the content of the recorded videos, varying veracity (true and false), and plausibility (plausible and implausible). The fourth and last factor consisted of the two confederates communicating in the video clips, ensuring that the findings would not be unique for one confederate.

The results reported below did not differ significantly between the two confederates that were used for the video-clips, the data was therefore collapsed between the confederates.

### **Materials**

The truthfulness prime consisted of 15 trait terms, such as reliable, honest, and truthful. The deceit prime also consisted of 15 trait terms, for example dishonest, unreliable, and false. The participants in the control condition were given 30 of the most common Swedish words, taken from the Kelly-list (Kilgariff et al., 2014; see Appendix A for the complete lists). The list of words for the experiment conditions contained 15 of the words that were part of the control condition list.

The two male confederates contributed with four videos each, one plausible truth, one implausible truth, one plausible lie, and one implausible lie. The time to prepare was intentionally kept short to mimic the conditions of someone needed to spontaneously think of an event. The videos were between 15 and 37 seconds in duration. The narratives were in Swedish to avoid confounding effects of language.

The following narratives are examples of included stimulus material. One plausible truth detailed a day-to-day routine. One implausible truth was about an unusual trip with a containership. One of the plausible lies detailed an experience with alcohol that ended up with the person being taken care of by police officers. One of the implausible lies were about a huge winning on a lottery-ticket.

### **Procedure**

The instrument was developed on the Qualtrics Survey Platform (<https://www.qualtrics.com>). The participants joined the experiment online. They completed a consent form and were randomized into the control, the truthfulness, or the deceit condition. The participants were advised to sit by a



computer, and if possible, to avoid using a tablet or a smart phone when participating. The first phase was the memorization task, i.e., the priming. The participants were asked to state the words out loud during this task to facilitate memorization. The words were displayed in a randomized order one by one, and each word was visible for three seconds.

In the second phase the participants were randomized to watch one out of the eight videos. The third phase presented the participants with the thought-listing procedure and the Likert-scale questions. The participants then did the memory test in the fourth and final phase of the experiment. The participants were presented with a list of 60 words from the Kelly-list in a randomized order and asked to make a choice between old and new. Old for words that had been present during the first phase of the experiment, and new for words that weren't present during that phase.

### **Measures and Statistical analysis**

The answers to the thought-listing question formed the main and first dependent variable. The answers were analyzed following the examples of Clare and Levine (2019), and Levine et al. (2021). A list was obtained from the Levine et al. (2021) experiment, which guided the search for words regarding deception or veracity. All answers were read and assessed by the author and one research assistant independently. Any mentions of words, both positive and negative, related to veracity, deception, trust, reliability, demeanor, plausibility, or believability resulted in the word being highlighted for further scrutiny. All answers were then re-examined by the author of this study. Any highlighted word was evaluated in its context. All words about deception or veracity resulted in a score, which was then divided by the total amount of typed in words by the participant resulting in a percentage of words relating to veracity assessments.

The following typed in thoughts exemplifies what was included: "Is it possible to nag one's way into the United States?", "The person is describing what most people would see as a very positive event. He does it however with a completely neutral, or a somewhat depressed tone. He does not seem to be positively affected by the event.". Two examples of not included thoughts was: "Insane luck", "A story on how you can influence and be successful persuading verbally". These and other comments were excluded given that they touched on other phenomenon's than deception or veracity.

The percentage scores represented a weighted measure allowing a comparison between the different conditions in the relative occurrence and frequency of veracity assessments. The percentage scores were subjected to a robust ANOVA in R (Wilcox, 2017; Wilcox & Schönbrodt, 2022).

The ratings of the confederate and the message content on the four dimensions of veracity, formed dependent variable two to five. A dimensions factor was created for the purpose of testing hypothesis five. The factor used veracity and plausibility creating four groups. Group one consisted of participants that watched plausible truths, group two was for plausible lies, group three was for implausible truths, and group four was for implausible lies (see Table 3 for descriptive statistics). The data did not meet the assumptions for (M)ANOVA. A robust MANOVA was implemented in R using the Brunner-Munzel's test (Brunner & Munzel, 2000; Karch, 2021). MANOVA was used to assess how the groups differed on combinations of the veracity dimensions. The measure of plausibility was treated separately in an explorational analyzes using a robust ANOVA (Wilcox, 2017; Wilcox & Schönbrodt, 2022) in R.

To assess accuracy and truth bias a new binary variable was created using the statement reliability data. The question put to the participant was to assess the reliability of the statement, from one (1: fully unreliable) to ten (10: fully reliable). A correct judgement of veracity occurred if the participant indicated a 1-5 for a lie or between 6-10 for a truth. Participants that had indicated a six or above on the ten-point Likert-scale were considered to believe in the statement.

A parallel multiple mediation model (Hayes, 2022) tested hypothesis six in assessing the mediating effects of plausibility and representativeness on the relationship between credibility and reliability.

The memory test served as a manipulation check and resulted in correct hits and false alarms. In line with Signal Detection Theory (Hautus et al., 2022), a participant's response was coded as a hit when correctly choosing old for a word that was present during the memorization in the first phase of the experiment. A false alarm was generated when old was chosen for a word that was not present during the memorization. The difference between z-transformed hit-rates and false alarm rates were used to calculate  $d'$  and  $c$  (Green & Swets, 1966). The  $d'$  scores functions as a sensitivity measure and represents the difference between the transformed hit-rates and false-alarm rates. The  $c$  scores functions as a response bias measure, in this study indicating the participants willingness to choose old. The  $c$  score is based on the sum of the z-transformed hit and false-alarm rates (Hautus et al., 2022). The  $d'$  scores were subjected to ANOVA in JASP (JASP Team, 2022). Percentages of hits for the trait terms that made up the priming in the two experimental conditions were also calculated.

## Results

The results are reported below in the order of the research hypotheses. Descriptive statistics are presented in Table 1 to 3.

The frequency of typed in thoughts pertaining to deception or veracity assessments (H1) was at a low level. Only 14.1% ( $n = 15$ ) out of the 106 participants in the experiment reported here typed in thoughts about deception or veracity. A third of these comments ( $n = 5$ ) were of a positive nature, and consequently only 9.4% of all the participants spontaneously made a negative veracity assessment of the narrative and/or the confederate.

The results regarding the second hypothesis (H2) showed an overall truth-bias of almost 69%, ranging from 54.83% for implausible content to 84.09% for plausible content (Table 1).

An overall accuracy rate for the experiment reported in this study was assessed to 61.3%, which ranged from 40.9% for lies to 75.8% for truthful content (Table 1).

Table 1 gives descriptive statistics showing percentage of participants believing in the content (truth-bias), percentage of participants correctly assessing truthful content as truthful and deceptive content as deceptive (accuracy rate), and keywords from the thought-listing procedure.

Hypothesis three predicted that implausible content independent of the actual veracity of the content would generate the highest frequency of deception or veracity relevant thoughts in the thought-listing procedure (H3). Hypothesis four predicted that semantic priming of words about deceit together with implausible content would increase the frequency of veracity assessments in comparison to no priming and the priming of words about truthfulness (H4).

The frequency of typed in thoughts on deception or veracity were 10% for the control condition, 13.15% for the truthfulness and 18.42% for the deceit condition. Most of the comments (86%,  $n = 13$ ) were about implausible content, and 26% ( $n = 4$ ) were regarding actual lies (see Table 1). The two comments on plausible content were of a positive nature. The frequency was highest in the deceit prime condition for implausible truths, while mentions for implausible lies in that condition came second. Participants in both the control condition and the truthfulness prime condition mentioned veracity to a lesser extent (see Table 1). When combining the deceit prime and implausible content (regardless of veracity), 31.81% of the participants mentioned thoughts about deception or veracity in the thought-listing phase. In the control and truthfulness prime condition, 11.8% respectively 17.39% of the participants that viewed implausible content mentioned thoughts pertaining to deception or veracity. An ANOVA on the percentage scores of the respective semantic priming condition was not significant,  $F(2, 103) = .211, p = .811, \omega^2 = .015$ . Hence, even though there

was a discernable difference in the frequencies of veracity assessments, the weighted scores did not significantly differentiate the groups based on the factor of priming.

The manipulation check showed a discernable difference between the three levels of the priming factor in terms of memory performance (Table 1). The  $d'$  scores were subjected to ANOVA with the priming condition as the fixed factor. There was a significant effect of condition on the  $d'$  scores,  $F(2, 103) = 8.438, p < .001, \omega^2 = .123$ . Planned contrasts showed that the experimental conditions significantly differed in comparison to the control condition,  $t(103) = 3.99, p < .001$  (one-tailed). The difference between the two experimental conditions was not significant,  $t(103) = .95, p = .345$  (one-tailed). Participants in the truthfulness prime condition remembered on average 78% of the words that constituted the prime. Participants in the deceit prime condition remembered on average 76% of the priming words.

**Table 1**  
Frequencies, accuracy, truth-bias, and SDT scores

Plausible Truth						Implausible Truth				Overall for the experiment				
Condition	n	Mentions	Keyword(s)	Accuracy	TB	n	Mentions	Keyword(s)	Accuracy	TB	n	Mentions	Accuracy	TB
Control	11	9	"Trustworthy"	100	100	8	12.5	"Implausible"	50	50	30	10	66	70
Truth	11	9	"Honest person"	100	100	12	25	"Unbelievable", "Is it possible", "Good actor"	58.33	58.33	38	13.5	63	71.05
Deceit	8	0	N/A	75	75	12	41.66	"Not his story", "Improbable", "Unreal"	66.66	66.66	38	18.9	55	65.7
Total	30	6	N/A	93.33	93.33	32	28.12	N/A	59.37	59.37	106	14.15	61.3	68.86
Plausible Lies						Implausible Lies				SDT scores				
Condition	n	Mentions	Keyword(s)	Accuracy	TB	n	Mentions	Keyword(s)	Accuracy	TB	d'	CI	c	CI
Control	2	0	N/A	0	100	9	0	N/A	55.55	44.44	1.2	.93, 1.53	.14	.0, .28
Truth	4	0	N/A	25	75	11	9	"Possibly not true"	45.45	54.54	1.7	1.55, 1.92	.14	.0, .28
Deceit	8	0	N/A	25	75	10	20	"Unbelievable", "Implausible"	50	50	1.9	1.66, 2.1	.18	.11, .42
Total	14	0	N/A	21.43	78.57	20	7.5	N/A	75	75				
Overall for Plausible					Overall for Implausible									
Condition	n	Mentions	Accuracy	TB	n	Mentions	Accuracy	TB						
Control	13	7.6	84.61	100	17	11.8	52.94	47.05						
Truth	15	6.6	80	93.33	23	17.39	52.17	56.52						
Deceit	16	0	50	75	22	31.81	59.09	59.09						
Total	44	4.5	70.45	84.09	62	16.12	54.83	54.83						
Overall for Truths					Overall for Lies									
Condition	n	Mentions	Accuracy	TB	n	Mentions	Accuracy	TB						
Control	19	10.52	78.94	78.94	11	9	45.45	54.54						
Truth	23	17.39	78.26	78.26	15	6.66	40	60						
Deceit	20	25	70	70	18	11.11	38.88	55.55						
Total	62	17	75.8	75.8	44	9	40.9	59.09						

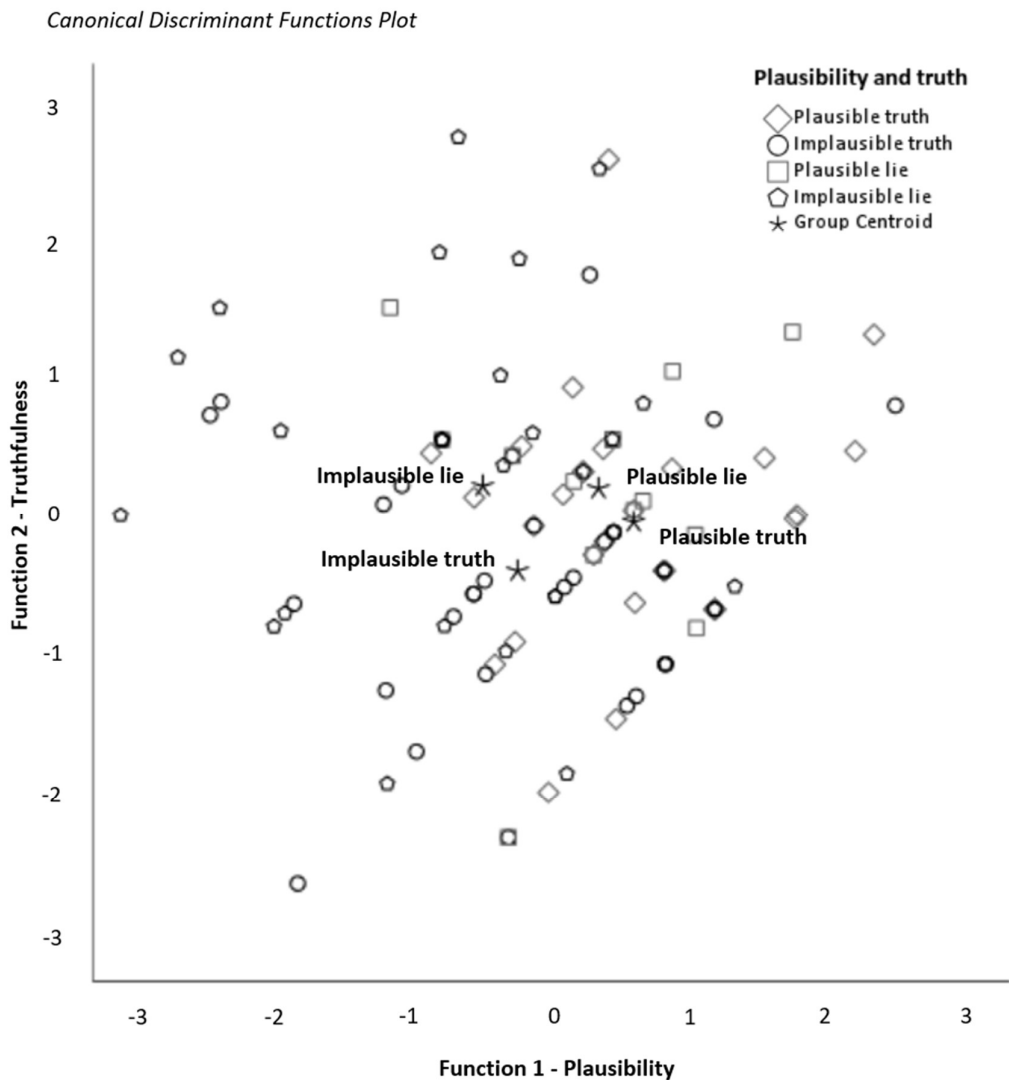
Note. Miniscule n gives sample size for the applicable condition. Mentions, accuracy, and truth-bias (TB) are given in percentage. Keyword(s) are representative samples of the typed in thoughts. The percentages are calculated for the number of participants stated under miniscule n. The Signal Detection Theory score  $d'$  indicates the average sensitivity of the participants in choosing old for old words in the memory test. A  $d'$  score of zero would be equal to performance at the level of chance. The  $c$  score indicates the average response bias. CI for SDT scores are given for the 95% level.

Message content plausibility was relative to message content veracity more important in differentiating the participants ratings of the communicator's credibility, statement reliability, representativeness, and content plausibility (H5).

A robust MANOVA (*mulrank* function; Wilcox, 2017) was used in R on the ranked data with the dimensions factor as the fixed factor. There was a significant main effect of plausibility and veracity on the outcomes of credibility, reliability and representativity,  $F = 7.73$ ,  $p < .001$ .

A discriminant function analysis revealed three discriminant functions. The first one explained 74.9% of the variance (canonical  $R^2 = 0.17$ ), whereas the second and third explained 24.2% (canonical  $R^2 = .06$ ) and .8% (canonical  $R^2 = .002$ ) respectively. In combination these discriminant functions significantly differentiated plausibility and veracity,  $\Lambda = .77$ ,  $\chi^2(9) = 26.62$ ,  $p = .002$ . When removing the first function from the analyzes the groups were not significantly differentiated,  $\Lambda = .933$ ,  $\chi^2(4) = 7$ ,  $p = .136$ . The canonical discriminant function plot (Figure 1) showed that the first function discriminated plausible from implausible, and the second discriminated truths from lies.

**Figure 1**



*Note.* This figure demonstrates how the different messages stimulus discriminated the groups along plausibility and truthfulness in the canonical discrimination analysis.

A robust ANOVA with dimensions factor as the fixed factor was done on the 20% trimmed means of the plausibility scores using the *t1way* function in R (Wilcox, 2017). There was a significant effect of truthfulness and content plausibility on the ratings of event plausibility,  $F(3, 101) = 18.9, p < .001, \omega^2 = .31$ . Planned contrasts using plausible truth as a baseline group showed significant differences, the comparison between plausible truth and plausible lie being the exception as not significant (Table 2).

**Table 2**

*Planned contrasts*

	Estimate	Std. error	<i>t</i>	<i>p</i>	<i>r</i> [95% CI]
Intercept	8.1667	.4501	18.146	<.001	
Implausible truth	-2.9731	.6313	-4.709	<.001	.54 [.33, .7]
Plausible lie	-.881	.7979	-1.104	0.272	.2 [-.1, .47]
Implausible lie	-4.233	.6365	-6.651	<.001	.67 [.51, .79]

*Note.* The effect size is calculated for the difference between means, in comparing each group with the plausible truth group. The *compute.es* function in R was used.

**Table 3***Ratings on the four dimensions of veracity*

Ratings of Trust on the Person								
<i>C</i>	<i>Mdn</i>	<i>M</i>	<i>CI (95%)</i>	<i>SD</i>	<i>Range</i>	<i>Mode</i>	<i>Skew</i>	<i>Kurtosis</i>
PT	8	8.13	7.52, 8.75	1.655	6	10	-.667	-.178
PL	7.5	7.36	6.28, 8.43	1.865	7	7, 8	-.776	1.22
IT	8	7.26	6.44, 8.07	2.221	8	8	-.680	.183
IL	6	5.97	4.96, 6.97	2.697	9	6	-.199	-.824

Ratings of Reliance on the Statement								
<i>C</i>	<i>Mdn</i>	<i>M</i>	<i>CI (95%)</i>	<i>SD</i>	<i>Range</i>	<i>Mode</i>	<i>Skew</i>	<i>Kurtosis</i>
PT	8	7.83	7.20, 8.46	1.683	6	10	-.277	-.688
PL	8	7.21	6.22, 8.21	1.718	6	8	-.492	-.576
IT	6	6	5.09, 6.91	2.49	9	5, 8	-.291	-.751
IL	5.5	5.6	4.66, 6.54	2.527	9	5	-.229	-.471

Ratings of Liar Representativeness								
<i>C</i>	<i>Mdn</i>	<i>M</i>	<i>CI (95%)</i>	<i>SD</i>	<i>Range</i>	<i>Mode</i>	<i>Skew</i>	<i>Kurtosis</i>
PT	4	4.27	3.24, 5.29	2.74	9	2, 5	0.594	.729
PL	5	4.86	3.82, 5.89	1.791	5	7	-.220	-1.073
IT	3	3.77	2.96, 4.59	2.217	9	5	.854	.729
IL	5	4.53	3.65, 5.42	2.374	8	5	.204	-.714

Ratings of Statement Plausibility								
<i>C</i>	<i>Mdn</i>	<i>M</i>	<i>CI (95%)</i>	<i>SD</i>	<i>Range</i>	<i>Mode</i>	<i>Skew</i>	<i>Kurtosis</i>
PT	8.5	8.17	7.46, 8.87	1.895	7	10	-1.201	1.026
PL	7.5	7.29	5.98, 8.6	2.268	9	6, 7, 8	-1.567	4.006
IT	5	5.19	4.19, 6.20	2.738	9	3	.273	-1.294
IL	3	3.93	2.91, 4.96	2.741	9	2	.839	-.676

*Note.* Results for the four different messages stimulus conditions. C stands for condition. PT: Plausible Truth. PL: Plausible Lie. IT: Implausible Truth. IL: Implausible Lie. Mdn gives the median rating. M gives the mean rating. CI and SD are given for the mean ratings. Range gives the total range of values for the given condition. Mode, skew and kurtosis describes the distribution.

The sixth hypothesis was about possible mediating effects between credibility and statement reliability through representativeness as a liar and/or message content plausibility (H6). TDT proposes that actual veracity is independent from honest-looking and deceptive-looking communicators. The mean ratings of credibility were all above the midpoint of 5.5 on the scale. Statement reliability ratings followed the same pattern.

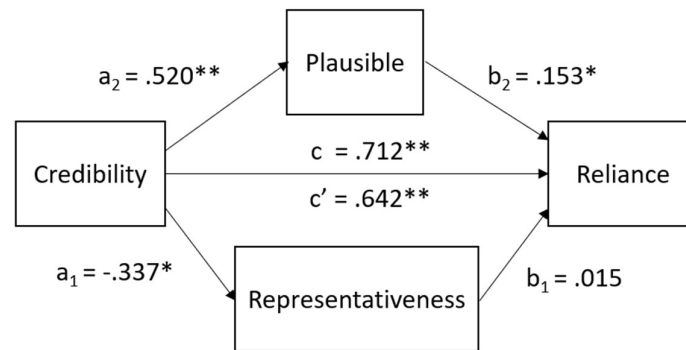
All means in the representativity ratings were below the midpoint of 5.5 on the scale. A low score was treated as indicative of behavior characteristic of someone telling the truth, while a high score was treated as indicative of someone behaving like a liar. The mean score for the participants viewing the plausible lie was the highest resulting in 4.86. The ratings of implausible truth resulted in 3.77, making it the lowest score.

The participants viewing implausible content were the only ones with means on the plausibility ratings below the midpoint of 5.5 on the scale (Table 3).

A parallel multiple mediator model with the scores from representativeness and plausibility as mediators revealed a significant indirect effect of credibility on reliance through plausibility,  $b_{a_2b_2} = .076$ , BC CI [.013, .17]. Credibility, as can be seen in Figure 2, was positively related to scores on plausibility but negatively related to representativeness. The relationship between credibility and reliance through representativeness was however not significant,  $b_{a_1b_1} = .005$ , BC CI [-.0534, .0486]. All bias-corrected confidence intervals were calculated for 95% confidence based on 5000 bootstrap samples (Hayes, 2022).

**Figure 2**

*The mediating effect of plausibility and representativeness*



*Note.* The mediating model demonstrates the paths from credibility through plausibility and representativeness to reliance. Miniscule a, b, c and c' gives the unstandardized regression coefficient.

\* =  $p < .05$ , \*\* =  $p < .001$

## Discussion

The aim of this study was to replicate the results presented by Clare and Levine (2019) and in part the results presented by Levine et al. (2021). This study also aimed to provide further insights into the effects of the truth-default state in processing another's communication. Results from previous experiments are reiterated when needed for the sake of clarity. The result from the experiment reported here are compared with those previous findings and discussed. This is followed by a general discussion on limitations of the current research, implications, and a concluding remark.

The first hypothesis predicted that between 5% and 20% of the participants would type in deception or veracity related thoughts.

The control condition of the experiment reported here resulted in 10% of the participants mentioning veracity. One of those mentions were in a positive regard. The deceit condition had the highest frequency of mentions resulting in 18.42% of the participants mentioning veracity (see also Table 1).

Both the Clare and Levine (2019) and Levine et al. (2021) studies showed strong support for the existence and effects of a truth-default state on assessing the veracity of communication. The results from the experiment reported here further corroborates those previous findings. To summarize, the results suggests that people tend to passively presume that a communicated narrative



is truthful and that senders are trustworthy. The fact that a demographically different sample replicates the earlier findings emphasizes that people rarely assess the veracity of a communicated message unless asked to do so. Hypothesis one was supported.

The second hypothesis was regarding the levels of truth-bias and accuracy. Clare and Levine (2019) reported that more than half (58%) of all participants in their first experiment believed in the contents of the messages. For the current experiment an overall truth-bias was assessed to 68.86%, and half of all participants in the implausible lie condition believed in the narratives. The truth-bias in the experiment reported in this study was on an overall level higher than the results presented by Clare and Levine (2019).

In the present study, the comments on the video clips provides an alternative angle in assessing the rate of believability. The majority (90.5%) of the participants did not use any negative words, i.e., implying that the narrative was deceitful, during the thought-listing phase. A small number of participants (4.7%) gave positive remarks about the confederate's credibility or about the reliability of the narrative.

Total accuracy for judging truths as truthful and lies as deceitful was at a level of just under 60% in the experiment reported by Clare and Levine (2019). The participants in the experiment reported here reached an overall accuracy level at 61.3%. Hence, the level of overall accuracy is replicated.

It can also be noted that 75.8% of the participants in the current experiment were accurate in assessing truthful messages as true. The level of accuracy for lies was 40.9%. Higher accuracy rates for truths than lies are a consequence of the veracity effect, one of the modules in TDT mentioned in the introduction. We presume that messages are true, and this results in people assessing most communication as truthful. Hence, when we assess veracity, we are truth-biased which results in more messages being believed than not believed. In summary, hypothesis two was supported by the results in the experiment done for this study.

Hypothesis three predicted that implausible content independent of actual veracity would generate the highest frequency of veracity assessments. Implausibility was in line with the results from Clare and Levine (2019) an important factor for spontaneously assessing veracity. Participants viewing an implausible narrative (regardless of veracity) were almost two times more likely to type in thoughts about deception or veracity than participants viewing a deceptive (regardless of plausibility) narrative (see Table 1).

This experiment had a skewed rate of implausible and truthful content (58.5%), which affected the frequencies of assessments for the different conditions of messages. Even when taking this skewed rate into account, 86% of the comments were about truthful implausible content. This can be compared to the 26% that commented on implausible lies. The plausible lie condition generated no comments. The comments on plausible truthful content were furthermore on a positive note.

The robust ANOVA on the plausibility scores underlined the relative importance of plausibility in comparison to veracity. There was only a minor difference when comparing plausible lies to plausible truths. Hence, participants did not make a discernable difference between these two types of message stimulus. However, plausible versus implausible content did make a significant difference. To summarize, the third hypothesis was supported by the results in the experiment reported in this study.

Hypothesis four predicted that semantic priming of words about deceit together with implausible content would increase the occurrence and frequency of veracity assessments in comparison to no priming and the priming of words about truthfulness

TDT proposes that a sufficiently potent trigger makes a person abandon the truth-default state and become suspicious (Levine, 2020). This experiment tested semantic priming as a triggering event. Both trait terms for truthfulness and trait terms for deceit did not substantially differentiate the groups on an overall experimental level in the weighted occurrence and frequency of veracity assessments. However, 31.81% of the participants mentioned veracity when combining the deceit

prime and implausible content. In the control and truthfulness prime condition, 11.8% respectively 17.39% of the participants that viewed implausible content mentioned veracity. A participant in the deceit prime condition was almost three times as likely to type in a veracity assessment in comparison to a participant in the control condition. TDT holds that a second trigger is needed to shift from suspicion to active disbelief (Levine, 2020; Levine et al., 2021). The priming of words about deceit coupled with implausible content increased the frequency of veracity assessment in comparison to the truthfulness priming condition and the control condition. This result supports TDTs view on the need of two triggers to abandon the truth-default state. The manipulation check showed a difference in terms of  $d'$  when comparing the two experimental conditions with the control condition. In addition, the percentage of hits for the trait terms were above 75% for both experimental conditions. Hence, the participants retained and recalled the trait terms to a large degree. To summarize, the results suggests that the priming as implemented in this study coupled with implausible content made a discernable difference on the percentage of participants typing in a veracity assessment. Hypothesis four was supported.

The fifth hypothesis predicted that message content plausibility rather than message content veracity would be the more important factor in differentiating the participants ratings of the communicator's credibility, statement reliability, representativeness, and content plausibility.

Most lies are discovered thanks to information from third parties, and most discoveries happens days, weeks, months, or even years after the lie occurred (Park et al., 2002; Masip & Herrero, 2015). Information from third parties plays an important role in the Strategic Use of Evidence approach in detecting deception (Granhag et al., 2007; Luke et al., 2018). The current study shows that plausible statements can get high ratings of plausibility independent of actual veracity.

The discriminant function analysis showed that differences of plausibility explained almost 75% of the variance in ratings of credibility, reliability in the statement and the ratings of the confederates' behavior as characteristic of someone lying. This can be compared to the result for difference in veracity, which explained 24% of the variance. Ratings were to a larger extent associated with differences in plausibility than actual veracity. The results in the experiment reported in this study suggests that humans are less sensitive to differences in veracity than plausibility. Humans appear to be truth-biased in the sense that we passively presume that others are honest. This is a harmless tendency in most day-to-day situations and TDT holds that efficient communication is facilitated through the assumption of honesty (Levine, 2020). The truth-default state is adaptive, and one can abandon the assumption of honesty given sufficiently potent triggers. Dynel (2011) sees implausibility as an overt violation of the Gricean maxim of quality, in line with reasoning by McCornack (1992). People in general have some knowledge about what is usual and/or typical given the context (Blair et al., 2010; Levine, 2020). Levine (2020) considers this knowledge as a possible trigger, detailing it as "a lack of correspondence between communication content and some knowledge of reality" (p. 99). Hence, implausibility can result in the receiver not relying on the message content to the same extent as if no Gricean maxim violation had occurred. Given this line of reasoning, it was predicted that content plausibility would take precedence over message content veracity for the veracity dimensions ratings. The fifth hypothesis was supported by the experimental results reported in this study.

The sixth hypothesis was that ratings of credibility and ratings of statement reliability is mediated by ratings of representativeness as a liar and of message content plausibility

TDT proposes that demeanor can be a confounding factor for making correct veracity assessments (Levine, 2014). The scores for the representativeness ratings didn't mediate the relationship between credibility and reliance, while plausibility did. The scale for representativeness was inverted in comparison to the other Likert-scale questions. A high score on credibility and reliance should generate a low score on representativeness. This was however not always the case, 15.1% of the participants with ratings of trust above a five on the scale also rated representativeness above a five on the scale. The ratings of statement reliability were also above five for 14.2% of the participants.

The mediation analysis showed a negative relationship between credibility and liar representativity. Thus, when a participant rated a confederate's behavior as characteristic of a liar, the participant generally also gave a low rate of credibility. The result in this experiment suggests however that the relationship between credibility and reliability cannot be explained by their relationship to representativeness as a liar. This however can be an effect of time-strapped participants.

The relationship between credibility and reliance can be explained by their relationship to message plausibility. One of the triggers identified in TDT is as noted earlier a discrepancy between known facts by the receiver and the content of the communicated narrative. Out of 62 participants that viewed an implausible message, 13 participants mentioned a veracity assessment in the thought-listing phase. Four of these mentions were about actual lies. The sixth and final hypothesis was supported in terms of ratings of plausibility.

The semantic priming aimed to test the concluding remark that "absent prompting, priming, or another trigger, the possibility that a message might be deceptive simply does not come to mind" (Clare & Levine, 2019, p. 302). Semantic priming as implemented in this study did not substantially differentiate participants in the two experimental conditions from the control condition on an overall level in the experiment. Priming coupled with implausible content differentiated the experimental conditions in the frequency of veracity assessments, supporting TDT's view on the need of two triggers to abandon the truth-default state. The differences between the conditions on the memory test result suggests that the words used for the priming were retained and easily accessible. However, the memory test result can be a product of the priming words being more easily distinguished and memorized in comparison to the more common words from the Kelly-list.

Doubts about the effects of priming, as noted in the introduction, has been raised given that some influential experiments have failed to replicate (Harris et al., 2013; but see Sherman & Rivers, 2021). The control condition in this study serves in any case as a reminder that spontaneous veracity assessments, absent prompting, priming, or a trigger are rare.

This study has a number of other limitations, one being the stimulus material. The material consisted of two males telling auto-biographical stories. It is possible that different results would have emerged if the narratives were about political or factual nature. RT, former Russia Today, produces misleading content relating to Russian culture among other topics (Elsawah & Howard, 2020). Wagnsson and Barzanje (2021) showed how state sponsored Sputnik News uses harmful strategic narratives in articles about Sweden. Both RT and Sputnik among others are in the business of producing junk news, which refers to "sources that deliberately publish misleading, deceptive, or incorrect information packaged as real news" (Bradshaw et al., 2020, p. 173; Howard, 2020). Future research could use fact-checked content from different junk news producers to assess the effects of the truth-default state when consuming that type of material.

A second limiting factor is the skewed rates of truths versus lies, and implausible versus plausible message content. Almost 60% of the content was of a truthful and implausible nature. Implausibility was an important factor behind veracity assessments, and the skewed rates most likely confounded the result. However, the overall truth-bias reached 68.86%. The lowest score for truth-bias was for implausible lies, but still 50% of those messages were believed. The results points towards higher levels of truth-bias than would be expected by the above-mentioned rates. It can also be noted that the base-rates of truths and lies outside the deception lab contains a higher proportion of truths than lies (Serota et al., 2010). Levine et al. (2014) varied the truth-lie base-rates in three experiments that used an interactive setting for the participants. When truthful content increased, a linear increase of accuracy in correctly assessing veracity followed (the before-mentioned veracity effect). This is in line with the Park-Levine probability model (Park & Levine, 2001), one of TDTs modules (Levine, 2020). Predictions and results have been replicated in Korea, showing the robustness and cross-cultural applicability of TDT (Park & Levine, 2017).

Implausible lies generated the highest frequency of veracity assessments in the Clare and Levine (2019) study. In this study implausible truths made up 60% of the typed in thoughts about deception or veracity. The skewed base-rate of truths and lies is noted as a possible confounding factor for this result. However, participants were almost two times more likely to spontaneously assess veracity on implausible truths than on implausible lies. In terms of assessing credibility and reliance on the statement, implausible lies resulted in the lowest mean scores. Future research could use prior research by Vrij et al. (2020) to see if an instruction to the participants to take note of frequency of details, complications, and verifiable sources would make a discernable difference in deception detection performance. Such an instruction to the participants in this study could have led to different results. Future research could also use the above-mentioned instruction coupled with different base-rates of truthful-deceitful and plausible-implausible content to research the effects on deception detection accuracy.

A third limiting factor is the indirect measure of truth-bias and accuracy. Different veracity dimensions were used in order not to confound the results by giving an additional trigger for participants to consider deception. Any indirect measure entails a degree of uncertainty. The participants in the experiment reported here were not allowed to make a dichotomous choice between truth or lie. The ratings of statement reliability were used instead of a dichotomous choice, given that an assessment about veracity is ultimately about the content of a communicated narrative. Future research could use different types of content and a dichotomous choice between truth and lie, such as the one employed by Clare and Levine (2019).

A fourth limiting factor is the fact that an absence of a veracity assessment was used to assess the presence of a truth-default state. Even though this assessment is also based on the different ratings the participants did, it is still an indirect measure. However, the fact that this study replicates the results from previous studies shows that a truth-default state indeed exists and makes us truth-biased.

This study began with a 20-year-old college student creating a conspiracy theory. The only thing more puzzling than the theory itself is the amount of people believing in it. Truth-default theory provides a framework to understand why people tend to believe what's being communicated to them.

The experiment sought to replicate and expand on the results from the Clare and Levine (2019) and Levine et al. (2021) studies. The results on levels of truth-bias, accuracy, and implausibility replicated in this study using a demographically different sample in a spatially and temporally different context. Thus, truth-default theory is supported.

Truth-bias and the truth-default state is harmless in most day-to-day communication given the low prevalence of lying (Serota & Levine, 2015; Serota et al., 2010; Daiku et al., 2021). Many of the lies told in everyday conversation are deemed necessary discursive elements (Turner et al., 1975). Turner et al. (1975) concluded that honesty isn't always the preferred option. Most day-to-day lies are seen as neither serious nor thoroughly planned (Turner et al., 1975; DePaulo et al., 1996).

There's however empirical evidence showing that the prevalence of lying is different on the internet in comparison with real life interactions (Drouin et al., 2016; Caspy & Gorsky, 2006). Voters in European elections are increasingly relying on social media for news and information on politics (Neudert et al., 2019). Social media as a source of political information has been linked to lower levels of trust for political institutions (Ceron, 2015). Social media has made it easier for communicators to reach larger audiences, becoming a platform for junk news, malign information influence and deceptive communicators. Malign information influence denotes "information sponsored by authoritarian regimes or other hostile actors and projected through international broadcasting to inflict harm upon others" (Wagnsson, 2022, p. 2).

The low frequency of veracity assessments in viewing content from YouTube highlights the fact that consumers of online content default to the truth and few thoughts about deception occurs

(Levine et al., 2021). Stanovich and West (2000) noted that conversational implicatures are in play when system 1 is used, even in situations devoid of conversational character. The result in this study supports that proposition. The results reported in this study and by Levine et al. (2021) shows that we are vulnerable in our consumption of online content where the prevalence of lying differs.

Malign information influence, junk news and people believing in “Birds Aren’t Real” poses a challenge to society. It highlights the need to educate in source criticism skills. The participants in the experiment reported here chose more often to trust the person in the online video clip and believe in the communicated content than the other way around. In conclusion, this study has provided additional insights on the robust nature of the truth-default state. People are truth-biased, a tendency that most of the time is harmless but can lead us to believe in malign information, junk news and deceptive communication.

## References

- Aamodt, M. G., & Custer, H. (2006). Who Can Best Catch a Liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, 15, 7-11.  
[https://www.researchgate.net/publication/232424344\\_Who\\_can\\_best\\_catch\\_a\\_liar\\_A\\_meta-analysis\\_of\\_individual\\_differences\\_in\\_detecting\\_deception](https://www.researchgate.net/publication/232424344_Who_can_best_catch_a_liar_A_meta-analysis_of_individual_differences_in_detecting_deception)
- AC Del Re. (2013). compute.es: Compute Effect Sizes. R package version 0.2-2.  
<https://cran.r-project.org/package=compute.es>
- Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In Uleman, J. S., and Bargh, J. A. (eds.), *Unintended Thought* (3–51). Guilford, New York.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action. *Journal of Personality and Social Psychology*, 71(2), 230-244. <https://doi.org/10.1037/0022-3514.71.2.230>
- Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36(2), 147-168. <https://doi.org/10.1002/ejsp.336>
- Bok, S. (1979). *Lying, Moral Choice in Public and Private Life* (1st ed). Vintage Books.
- Bond, C. F., and DePaulo B. M. (2006). Accuracy of Deception Judgements. *Personality and Social Psychology Review*, 10(3), 214-234. [https://doi.org/10.1207/s15327957pspr1003\\_2](https://doi.org/10.1207/s15327957pspr1003_2)
- Bond, C. F., and DePaulo B. M. (2008). Individual Differences in Judging Deception: Accuracy and Bias. *Psychological Bulletin*, 134(4), 477-492. <https://doi.org/10.1037/0033-2909.134.4.477>
- Bond, C. F., Levine, T. R., & Hartwig, M. (2015). New Findings in Non-Verbal Lie Detection. In Granhag, P. A., Vrij, A., & Verchueren, B. (Ed.), *Detecting Deception – Current Challenges and Cognitive Approaches* (p. 37-58). Wiley Blackwell.
- Bradshaw, S., Howard, P. N., Kollanyi, B., & Neudert, L-M. (2020). Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018, *Political Communication*, 37(2), 173-193. <https://doi.org/10.1080/10584609.2019.1663322>
- Brunner, E., & Munzel, U. (2000). The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal*, 42(1), 17-25.  
[https://dx.doi.org/10.1002/\(SICI\)1521-4036\(200001\)42:1%3C17::AID-BIMJ17%3E3.0.CO;2-U](https://dx.doi.org/10.1002/(SICI)1521-4036(200001)42:1%3C17::AID-BIMJ17%3E3.0.CO;2-U)
- Ceron, A. (2015). Internet, News, and Political Trust: The Difference Between Social Media and Online Media Outlets. *Journal of Computer-Mediated Communication*, 20(5), 487-503.  
<https://dx.doi.org/10.1111/jcc4.12129>
- Clare, D. D., & Levine, T. R. (2019). Documenting the Truth-Default: The Low Frequency of Spontaneous Unprompted Veracity Assessments in Deception Detection. *Human Communication Research*, 45(3), 286-308. <https://doi.org/10.1093/hcr/hqz001>

- Daiku, Y., Serota, K. B., & Levine, T. R. (2021). A Few prolific Liars in Japan: Replication and the Effects of Dark Triad Personality Traits. *PLoS ONE* 16(4): e0249815.  
<https://doi.org/10.1371/journal.pone.0249815>
- 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(5), 979-995.
- DePaulo, B. M., Lindsay, J.J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129 (1), 74-118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Drouin, M., Miller, D., Wehle, S. M. J., & Hernandez, E. (2016). Why Do People Lie Online? “Because Everyone Lies on the Internet”. *Computers in Human Behavior*, 64, 134-142.  
<https://dx.doi.org/10.1016/j.chb.2016.06.052>
- Dutilh Novaes, C., & Veluwenkamp, H. (2017). Reasoning Biases, Non-Monotonic Logics and Belief Revision. *Theoria*, 83(1), 29-52. <https://dx.doi.org/10.1111/theo.12108>
- Dynel, M. (2011). A Web of Deceit: A Neo-Gricean View on Types of Verbal Deception. *International Review of Pragmatics*, 3(2), 139-167. <http://dx.doi.org/10.1163/187731011X597497>
- Elsawah, M., & Howard, P. N. (2020). “Anything That Causes Chaos”: The Organizational Behavior of Russia Today (RT). *Journal of Communication*, 70(5), 623-645.  
<https://doi.org/10.1093/joc/jqaa027>
- Fallis, D. (2012). Lying as a Violation of Grice’s First Maxim of Quality. *Dialectica*, 66(4), 563-581).  
<https://dx.doi.org/10.1111/1746-8361.12007>
- Feeley, T. H., & Young, M. (1998). Humans as Lie Detectors: Some More Second Thoughts. *Communication Quarterly*, 46(2), 109-126. <https://dx.doi.org/10.1080/01463379809370090>
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601-613.  
<https://doi.org/10.1037/0022-3514.59.4.601>
- Gilbert, D. T. (1991). How Mental Systems Believe. (1991). *The American Psychologist*, 46(2), 107-119). <https://dx.doi.org/10.1037/0003-066X.46.2.107>
- Goffman, E. (1956). The Nature of Deference and Demeanor. *American Anthropologist*, 58(3), 473-502. <https://dx.doi.org/10.1525/aa.1956.58.3.02a00070>
- Granhag, P-A., Strömwall, L., & Hartwig, M. (2007). The SUE-Technique: The way To Interview to Detect Deception. *Forensic Update*, 88, 25-29.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley
- Grice, P. (1989). *Studies in the Way of Words* (paperback edition). Harvard University Press.
- Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being Honest about Dishonesty: Correlating Self-Reports and Actual Lying. *Human Communication Research*, 40(1), 54-72.  
<https://dx.doi.org/10.1111/hcre.12019>

- Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two Failures to Replicate High-Performance-Goal Priming Effects. *PLoS one*, 8(8), <https://doi.org/10.1371/journal.pone.0072467>
- Hartwig, M., & Bond, C. F. (2011). Why Do Lie-Catchers Fail? A Lens Model Meta-Analysis of Human Lie Judgements. *Psychological Bulletin*, 137(4), 643-659. <https://dx.doi.org/10.1037/a0023589>
- Hartwig, M., Granhag, P-A., Strömwall, L., & Kronkvist, O. (2006). Strategic Use of Evidence During Police Interrogations: When Training to Detect Deception Works. *Law and Human Behavior*, 30(5), 603-619. <https://dx.doi.org/10.1007/s10979-006-9053-9>
- Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2022). *Detection Theory: A User's Guide* (3<sup>rd</sup> ed.). Taylor & Francis
- Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3<sup>rd</sup> ed). The Guilford Press.
- Higgins, E. T., & Brendl, C. M. (1995). Accessibility and applicability: some "activation rules" influencing judgment. *Journal of Experimental Social Psychology*, 31(3), 218-243. <https://dx.doi.org/10.1006/jesp.1995.1011>
- Higgins, E. T., & Eitam, B. (2014). Priming... Shimming: it's about knowing when and why stimulated memory representations become active. *Social Cognition*, 32(Supplement), 225-242. <https://doi.org/10.1521/soco.2014.32.supp.225>
- Howard, P. (2020). *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. Yale University Press.
- JASP Team. (2022). JASP (Version 0.16.2)[Computer software].
- Kahneman, D. (2003). A Perspective on Judgement and Choice: Mapping Bounded Rationality. *American Psychologist*, 58(9), 697-720. <https://dx.doi.org/10.1037/0003-066X.58.9.697>
- Karch, J. D. (2021). Psychologist Should Use Brunner-Munzel's Instead of Mann-Whitney's U Test as the default Nonparametric Procedure. *Advances in Methods and Practices in Psychological Science*, 4(2), 1-14. <https://dx.doi.org/10.1177/2515245921999602>
- Kilgarriff, Adam; Charalabopoulou, Frieda; Gavrilidou, Maria; Johannessen, Janne Bondi; Khalil, Saoussan; Kokkinakis, Sofie Johansson; Lew, Robert; Sharoff, Serge; Vadlapudi, Ravikiran & Volodina, Elena. 2014. Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*, 48, 121-163, <https://doi.org/10.1007/s10579-013-9251-2>
- Levine, T. R. (2010). A Few Transparent Liars Explaining 54% Accuracy in Deception Detection Experiments, *Annals of the International Communication Association*, 34(1), 41-61. <https://doi.org/10.1080/23808985.2010.11679095>



- Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, 33(4), 378-392. <https://doi.org/10.1177/0261927X14535916>
- Levine, T. R. (2018). Ecological Validity and Deception Detection Research Design. *Communication Methods and Measures*, 12(1), 45-54. <https://doi.org/10.1080/19312458.2017.1411471>
- Levine, T. R., Kim, R. K., & Blair, J. P. (2010)b. (In)accuracy at Detecting True and False Confessions and Denials: An Initial Test of a Projected Motive Model of Veracity Judgements. *Human Communication Research* 36(1), 82-102. <https://doi.org/10.1111/j.1468-2958.2009.01369.x>
- Levine, T. R., Kim, R. K., & Hamel, L. M. (2010)a. People Lie for a Reason: Three Experiments Documenting the Principle of Veracity. *Communication Research Reports*, 27(4), 271-285. <https://doi.org/10.1080/08824096.2010.496334>
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect.” *Communication Monographs*, 66(2), 125-144. <https://doi.org/10.1080/03637759909376468>
- Levine, T. R., Punyanunt-Carter, N. M., & Moore, Alivia. (2021). The Truth-Default and Video Clips. Testing the Limits of Credulity. *Human Communication Research*, 40(4), 442-462. <https://doi.org/10.1080/10510974.2020.1833357>
- Levine, T. R., Shaw, A., & Shulman, H. (2010). Increasing deception detection accuracy with strategic questioning. *Human Communication Research*, 36(2), 216-231. <https://doi.org/10.1111/j.1468-2958.2010.01374.x>
- Lloyd’s Register Foundation & Gallup (2019). *The Lloyd’s Register Foundation World Risk Poll. Full Report and Analysis of the 2019 Poll.* [https://wrp.lrfoundation.org.uk/LRF\\_WorldRiskReport\\_Book.pdf](https://wrp.lrfoundation.org.uk/LRF_WorldRiskReport_Book.pdf)
- Lorenz, T. (Dec. 9, 2021). Birds Aren’t Real, Or Are They? Inside a Gen Z Conspiracy Theory. *New York Times*. <https://www.nytimes.com/2021/12/09/technology/birds-arent-real-gen-z-misinformation.html>
- Luke, T. J. (2020). A Meta-Analytic Review of Experimental Tests of the Interrogation Technique of Hans Joachim Scharff. *Applied Cognitive Psychology*, 35(2), 360-373. <https://dx.doi.org/10.1002/acp.3771>
- Luke, T. J., Hartwig, M., Joseph, E., Brimbal, L., Chan, G., Dawson, E., Jordan, S., Donovan, P., & Granhag, P-A. (2016). Training in the Strategic Use of Evidence Technique: Improving Deception Detection Accuracy of American Law Enforcement Officers. *Journal of Police and Criminal Psychology*, 31(4), 270-278. <https://dx.doi.org/10.1007/s11896-015-9187-0>
- Luke, T. J., Hartwig, M., Brimbal, L., Granhag, P-A. (2018). Building a Case: The Role of Empirically Based Interviewing Techniques in Case Construction. In H. Otgaar & M. L. Howe (Ed.), *Finding the Truth in the Courtroom*. Oxford University Press.
- Mandelbaum, E. (2014). Thinking is Believing. *Inquiry*, 57(1), 55-96. <https://doi.org/10.1080/0020174X.2014.858417>

- McCornack, S. A. (1992). Information Manipulation Theory. *Communication Monographs*, 59(1), 1-16. <https://doi.org/10.1080/03637759209376245>
- McCornack, S. A., & Parks, M. R. (1986). Deception Detection and Relationship Development: The Other Side of Trust. In M. L. McLaughlin (Ed.), *Communication Yearbook 9* (377-389). Sage. <http://dx.doi.org/10.1080/23808985.1986.11678616>
- Meijer, E. H., & Verschuere, B. (2018). Detection Deception Using Psychophysiological and Neural Measures. In H. Otgaar & M. L. Howe (Ed.), *Finding the Truth in the Courtroom*. Oxford University Press.
- Mussweiler, T., & Strack, F. Hypothesis-Consistent Testing and Semantic Priming in the Anchoring Paradigm: A Selective Accessibility Model. *Journal of Experimental Social Psychology*, 35(2), 136-164. <https://dx.doi.org/10.1006/jesp.1998.1364>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology*, 69, 511-534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Park, H. S., & Levine, T. R. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs*, 68(2), 201-210. <https://doi.org/10.1080/03637750128059>
- Park, H. S., & Levine, T. R. (2017). The Effects of Truth-Lie Base-Rates on Deception Detection Accuracy in Korea. *Asian Journal of Communication*, 27(5), 554-562. <https://doi.org/10.1080/01292986.2017.1334074>
- Park, H. S., Levine, T., McCornack, S., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, 69(2), 144-157. <https://doi.org/10.1080/714041710>
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2017). Bayes Factor Design Analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128-142. <https://dx.doi.org/10.3758/s13423-017-1230-y>
- Sherman, J. W., & Rivers, A. M. (2021). There's Nothing Social about Social Priming: Derailing the "Train Wreck". *Psychological Inquiry*, 32(1), 1-11. <https://doi.org/10.1080/1047840X.2021.1889312>
- Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The Prevalence of Lying in America: Three Studies of Self-Reported Lies. *Human Communication Research*, 36(1), 2-25. <https://doi.org/10.1111/j.1468-2958.2009.01366.x>
- Serota, K. B., & Levine, T. R. (2015). A Few Prolific Liars: Variation in the Prevalence of Lying. *Journal of Language and Social Psychology*, 34(2), 38-157. <https://dx.doi.org/10.1177/0261927X14528804>
- Sorensen, R. (2007). Bald-Faced Lies! Lying Without the Intent to Deceive. *Pacific Philosophical Quarterly*, 88(2), 251-264. <https://dx.doi.org/10.1111/j.1468-0114.2007.00290.x>

- Sporer, L. S., & Schwandt, B. (2007). Moderators of Nonverbal Indicators of Deception: A Meta-Analytic Synthesis. *Psychology, Public Policy, and Law*, 13(1), 1-34.  
<https://doi.org/10.1037/1076-8971.13.1.1>
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning. Implications for the rationality debate. *The Behavioral and Brain Sciences*, 23(5), 645-665.  
<https://dx.doi.org/10.1017/S0140525X00003435>
- Turner, R. E., Edgley, C., & Olmstead, G. (1975). Information Control in Conversations: Honesty Is Not Always the Best Policy. *The Kansas Journal of Sociology*, 11(1), 69-89.  
<http://www.jstor.org/stable/23255229>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131. <https://www.jstor.org/stable/1738360>
- Vrij, A. (2000). *Detecting Lies and Deceit: the psychology of lying and the implications for professional practice* (2nd ed.). Wiley.
- Vrij, A., Deeb, H., Leal, S., Granhag, P-A., Fisher, R. P. Plausibility: A Verbal Cue to Veracity Worth Examining? (2020). *The European Journal of Psychology Applied to Legal Context*, 13(2), 47-53. <https://doi.org/10.5093/ejpalc2021a4>
- Wagnsson, C. (2022). The Paperboys of Russian Messaging: RT/Sputnik Audiences as Vehicles for Malign Information Influence. *Information, Communication, & Society* (ahead of print).  
<https://doi.org/10.1080/1369118X.2022.2041700>
- Wagnsson, C., & Barzanje, C. (2021). A Framework for Analysing Antagonistic Narrative Strategies: A Russian Tale of Swedish Decline. *Media, War, and Conflict*, 14(2), 239-257.  
<https://dx.doi.org/10.1177/1750635219884343>
- Was, C., Wolz, D., & Hirsch, D. (2019). Memory processes underlying long-term semantic priming. *Memory and Cognition*, 47(2), 313-325. <https://dx.doi.org/10.3758/s13421-018-0867-8>
- Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From Primed Concepts to Action: A Meta-Analysis of the Behavioral Effects of Incidentally Presented Words. *Psychological Bulletin*, 142(5), 472-497. <https://dx.doi.org/10.1037/bul0000030>
- Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (Fourth edition). Academic Press.
- Wilcox, R. R., & Schönbrodt, F. D. (2022). The WRS package for robust statistics in R (version 0.39).  
<https://github.com/nicebread/WRS>
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. *Advances In Experimental Social Psychology*, 14, 1-59.  
[https://doi.org/10.1016/S0065-2601\(08\)60369-X](https://doi.org/10.1016/S0065-2601(08)60369-X)

## Appendix A – The memorized words

Control		Positive		Negative	
Swedish	English	Swedish	English	Swedish	English
analog	analogue	<b>ärlig</b>	<b>honest</b>	dålig	bad
bra	good	<b>autentisk</b>	<b>authentic</b>	december	december
dålig	bad	dålig	bad	ekonomisk	economical
december	december	december	december	enda	only
egen	own	ekonomisk	economical	<b>falsk</b>	<b>false</b>
ekonomisk	economical	enda	only	flest	most
enda	only	flest	most	<b>förtegen</b>	<b>reticent</b>
fler	more	god	tasty	god	tasty
flest	most	intressant	interesting	<b>illojal</b>	<b>disloyal</b>
gammal	old	januari	january	<b>inkorrekt</b>	<b>incorrect</b>
god	tasty	<b>korrekt</b>	<b>correct</b>	intressant	interesting
hel	whole	lång	long	januari	january
hög	high	<b>lojal</b>	<b>loyal</b>	lång	long
intressant	interesting	<b>övertygande</b>	<b>persuasive</b>	<b>lögnaktigt</b>	<b>deceitful</b>
januari	january	<b>pålitlig</b>	<b>reliable</b>	<b>oärlig</b>	<b>dishonest</b>
lång	long	politisk	political	<b>opålitlig</b>	<b>unreliable</b>
lik	alike	<b>reell</b>	<b>real</b>	<b>orimlig</b>	<b>implausible</b>
mycket	a lot	<b>riktig</b>	<b>proper</b>	<b>osant</b>	<b>untrue</b>
ny	new	<b>rimlig</b>	<b>reasonable</b>	<b>otrogen</b>	<b>unfaithful</b>
olik	different	<b>sanningsenlig</b>	<b>truthful</b>	<b>otroligt</b>	<b>incredible</b>
politisk	political	<b>sant</b>	<b>true</b>	<b>overklig</b>	<b>unreal</b>
sen	late	sen	late	<b>påhittat</b>	<b>made up</b>
sista	last	sista	last	politisk	political
stor	big	svår	difficult	sen	late
svår	difficult	<b>trogen</b>	<b>faithful</b>	sista	last
svensk	Swedish	<b>trolig</b>	<b>plausible</b>	svår	difficult
ung	young	<b> trovärdig</b>	<b>believable</b>	<b>tveksam</b>	<b>doubtful</b>
vanlig	ordinary	ung	young	<b>tvivelaktig</b>	<b>dubious</b>
viktig	important	<b>uppriktig</b>	<b>sincere</b>	ung	young
viss	certain	vanlig	ordinary	vanlig	ordinary

Notes. The words used for priming are marked in bold.