Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

~~Review~~

Truth-default theory and the psychology of lying and deception detection

Timothy R. Levine

Abstract

Truth-default theory offers an account of human deceptive communication where people are honest unless they have a motive to deceive and people passively believe others unless suspicion and doubt are actively triggered. The theory is argued to account for wide swings in vulnerability to deception in different types of situations in and out of the lab. Three moderators are advanced to account for differential vulnerability to political misinformation and disinformation. Own belief congruity, social congruence, and message repetition are argued to combine to affect the probability that implausible and refutable false information is accepted as true.

Addresses

University of Alabama at Birmingham, Heritage Hall, Room 302, 1401 University Blvd., Birmingham, AL 35233, ~~England~~

Corresponding author: Levine, Timothy R. (levinet111@gmail.com)

Current Opinion in Psychology 2022, **47:**101380

This review comes from a themed issue on **Honesty and Deception**

Edited by **Maurice E. Schweitzer** and **Emma Levine**

For a complete overview see the Issue and the Editorial

Available online 3 June 2022

https://doi.org/10.1016/j.copsyc.2022.101380

2352-250X/© 2022 Elsevier Ltd. All rights reserved.

Keywords

Misinformation, Disinformation, Confirmation bias, Social proof, Lies.

> "*The truth has no defense against a fool determined to believe a lie.*" Apparently fake Mark Twain quote posted on twitter by Ref. [1].

Introduction

The twin goals of this essay are to introduce readers to truth-default theory [2,3] and to expand TDT by specifying three new trigger moderators. These additions allow TDT to make better sense of human vulnerability to implausible and refutable false information. By implausible lies, I mean communication content that should be obviously false such as the claim that birds aren't real [4] and other seemingly improbable claims.

My argument begins with three examples demonstrating different levels of vulnerability to false or misleading information: fraudulent text messages, research involving deceptive research practices, and research explicitly testing deception detection accuracy. Later in the essay, the case of false political information is added to the mix. My ultimate aim is to show how a modified TDT provides a coherent and robust account of deception in and out of the lab.

Explaining deception detection in everyday life and the lab

Three examples

I typically get several deceptive text messages per day. They are easy to spot. For example, I was just informed that my $9377 Winter stimulus benefit was just issued. All I need to do is to reply "Winter" to claim it. I hit "report spam" instead and blocked the number.

Most people, I think, are quite good at detecting deception like this. I estimate that the accuracy of deception detection for fraudulent text messaging is probably better than 99.9%. With an estimated 87.8 billion spam text messages per year [5], only a tiny proportion need to be successful to defraud millions from victims. If the projected loss in 2020 was $86 million and the median loss was $800 [17], then the hit rate is low indeed. The point is that the vast majority of deceptive text messages do not fool their targets.

Compare deception detection in fraudulent text messaging to experiments that involve stooges or confederates. These are examples of identity deception where the people the research participants encounter are not who they claim to be. In my lab, these are members of my research team who pose as research participants to stage an experiment. Famous examples include Ref. [6] group members and Ref. [7] victim. As those who do this sort of research know well, research participants seldom suspect that others in the experiments are not who they seem. I estimate that accuracy in this sort of situation hovers around 1% or 2%. In one of my recent experiments, only 1 out of 67 participants presented even minimal evidence of suspicion, and that study included an implausible lie condition on top of identity deception [8]. All of the participants unquestionably accepted that they were conversing with another research participant. They were not. The deception was invariably successful. Accuracy was 0.00%

The consequences of being duped by a criminal scam compared to an IRB-approved experiment are quite different. The severity of consequences, however, is not why these situations produce such different levels of accuracy. Before getting to my explanation, there is a third example to ponder.

If you read peer-reviewed, social scientific research on deception detection, it won't take you long to come across the 54% accuracy finding. It is ubiquitous in the literature. The 54% finding comes from the often-cited and authoritative meta-analysis by Ref. [9]. Accuracy (raw percent correct truth-lie discrimination) was normally distributed around 54% with a standard deviation of about 6%. Hit rates in deception detection experiments are uniformly much lower than for fraudulent text messages and much higher than for identity deception in lab experiments.

A puzzle for theories of human deception detection

Why, I ask, is deception detection accuracy better than 99% for fraudulent text messages, 54% in direct experimental tests, and less than 2% in a different experimental literature involving deception but not about deception? TDT, I argue, explains these large variations.

Truth-default theory

As the name of the theory implies, the central most idea of TDT is that people default to the truth, or at least truth as they know it. Generally, humans communicate honestly unless there is a reason(s) to deceive others, and humans passively accept incoming communication as honest unless they have a reason(s) to suspect otherwise. That is, deception (but not honest communication) requires a motive. Suspicion, skepticism, and disbelief require triggers. As both senders and receivers of communication, honesty and passive belief are the starting points.

The focus of this essay is on the reception side of TDT. From the TDT perspective, defaulting to the truth is highly beneficial. It allows for effective and efficient communication. We can learn from other people and pass along our knowledge. We get the benefits of cooperation and coordinating our activities with others. We develop and maintain social, personal, and professional relationships. We would be bogged down in uncertainty if we second-guessed the veracity of all incoming communication. Further, most communication is honest, and even when deceptive, most lies are benign [10]. The risks of harm from deception does not justify constant vigilance. The cost of defaulting to the truth is short-term vulnerability to occasional deception. In the TDT view, the trade-off is more than worth it [3].

This said, people can be too trusting or too prone to suspicion. People who miss red flags for deception are, for example, at a higher risk of victimization from financial fraud. Alternatively, overly suspicious people may have difficulty forming satisfying relationships. Situational calibration is critical.

According to TDT, the thought that any communication message might be deception does not come to mind unless actively triggered. TDT specifies two types of triggers. Triggers of the first kind – suspicion or skepticism triggers – bring the possibility of deception to mind. Triggers of the second kind move the cognitive state from suspicion and uncertainty to active disbelief. Examples of triggers include discrepancies between communication content and prior knowledge, logical inconsistencies in communication content, apparent motives for deception, and sender behaviors associated with a dishonest demeanor.

The two triggers are thought of as having different thresholds for activation or different levels of trigger sensitivity. When the first trigger is engaged but not the second, people are said to be truth-biased, although they can be truth-biased or suspicious in varying degrees. Unlike defaulting to the truth which is passive, people often will consciously accept communication as honest even when they are aware that it might not be.

It follows that for people to be accurate at deception detection, they first need to be on guard for it. The threshold for the first trigger must be passed, and people need to temporarily abandon their truth default state. Then, there must be diagnostically useful information in the communication and/or the environment to activate the second trigger and lead to a correct conscious assessment that the communication is false or misleading.

These two criteria seem to be met in the text fraud situation. People know that spam and phishing are things about which they need to be wary. Most of the fraud messages are variations on a set of themes and thus recognizable for what they are. Hence deception is accuracy high.

In the deceptive research practices and other instances of identity deception, the idea that things are not what they seem and people are not who they claim to be may never come to mind (Levine et al., 2020). The first trigger is not activated, and even if it is, the lab set up is such that the nature of deception is disguised. Participants may know that something might be hidden, but they do not know what is hidden. The thresholds for both triggers are not passed, and the participants are almost always duped.

In deception detection experiments, researchers activate the first trigger by explicitly asking the participants

Predictions about deception detection in various situations

Situation	Ratio honest and deceptive messaging	Suspicion Trigger	Availability of diagnostic veracity information	Anticipated accuracy	Moderators
Deception detection experiments	50–50	Yes	No	54%	Variables that increases diagnostic veracity information
Identity deception in experiment	100% deception	No	No	Less than 2%	No strong moderators
Fraudulent text messages	Variable	Yes	Yes	Greater than 99%	Knowledge of phishing
Political misinformation and disinformation	Variable	Variable	Yes	Highly variable	Belief congruence, social congruence, and message prevalence

to make a truth-lie assessment. Truths and lies are equally probable in the lab, and most senders do not give off diagnostic cues. Some senders are matched (come off as they are) and some are mismatched leading to predictable errors. A few people, however, are poor liars and are thus easily detectable keeping accuracy just above pure chance. The net result is 54% in deception detection experiments [3].

The case of misinformation/disinformation

Whereas TDT explains the differential detection rates in the previous examples, the original specification seems inadequate to account for another class of false or misleading content: misinformation, disinformation, fake news, and conspiracy theories [16]. Whether Q Anon acceptance, climate change denial, unjustified vaccine skepticism, or acceptance of the claims advanced by a particularly truth-challenged political figure, understanding the uncritical acceptance of false or implausible communication content is not as simple as merely defaulting to the truth. The problem for TDT in examples of this type is that triggers abound, but reasonable skepticism and justifiable disbelief are not sufficiently triggered and diagnostic information is dismissed. The theoretical solution, I believe, is the specification of trigger sensitivity moderators. I think that three, especially in combination, are most critical.

Before outlining the three moderators, I want to assert that in a theoretically important respect, the current misinformation/disinformation crisis in not new. Social and digital media amplify messaging and allow for rapid spread as never before. Nevertheless, large groups of humans have probably always held sets of beliefs that other groups dismiss as total nonsense. Human cultures have always had their mythologies. There is, I believe, something very deeply human about uncritically accepting objectively false (or at least unprovable) ideas.

Three moderators

Why, and under what conditions, will people accept information in the face of good reason to think it may be false? The first and most critical consideration I think is belief congruence. People are much more likely to believe something when it fits with what they already believe. People are also more likely to reject opinions and evidence that conflict with their beliefs. The more committed people are to their relevant beliefs, the stronger the effect. We can think of this as an example of confirmation bias [11] and understand it as an application of dissonance theory [12].

The second moderator is social congruence. What I have in mind here is a combination of [13] principles of social poof and identity. People are more likely to accept message content that is supported by their social network, especially when acceptance and rejection serves to differentiate an in-group from an outgroup.

The third moderator is message repetition and persistence. The more often people encounter the content, the more likely they are to accept it [14] and the more comfortable they become with the content [15].

My contention is when people repeatedly encounter message content that fits with their existing beliefs, matches what others in their social group believe, where the content is tied to an ingroup-outgroup distinction, and where the claim is ubiquitous in their communication environment, acceptance is highly probable regardless of message veracity or the availability of conflicting information. Under the confluence of these three conditions, people are almost invariably duped by false or misleading content.

Conclusion

Truth-default theory proposes that people default to the truth. Honesty is the starting point for both senders and receivers of communication. Humans communicate honestly unless there is a reason to deceive others, and suspicion, skepticism, and disbelief of communication require a trigger. A puzzle, however, is why people accept some types of false and misleading information even when there exist good, objective reasons to suspect veracity. The answer is that the relations between the content of the information, people's own beliefs, people's social environment, and the larger information environment combine to affect susceptibility to false and misleading information.

Credit author statement

Timothy R. Levine is the sole author of this essay and responsible for the conceptualization and writing. Dr. Dave Markowitz provided helpful comments on an early draft of this essay and Dr. Maurice Schweitzer provided feedback on a subsequent draft.

Conflict of interest statement

The author has no conflicts of interest to report.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

1. @Key3Skeleton: ***Twitter post***. March 17, https://twitter.com/key3skeleton/status/1372055234164441089?lang=en.

2. Levine TR: **Truth-default theory (TDT): a theory of human deception and deception detection**. *J Lang Soc Psychol* 2014, **33**:378–392.

3. * Levine TR: *Duped: truth-default theory and the social science of lying and deception*. Tuscaloosa, AL: University of Alabama Press; 2020.

This book provides a recent review of the literature on deception, explicates truth-default theory, and reports 55 studies and experiments testing truth-default theory.

4. Lorenz T: *Birds aren't real, or are they? Inside a gen z conspiracy theory*. The New York Times; 2021, December 9. https://www.nytimes.com/2021/12/09/technology/birds-arent-real-gen-z-misinformation.html.

5. RoboKiller: ***RoboKiller releases 2021 phone scam report***. https://www.prnewswire.com/news-releases/robokiller-releases-2021-phone-scam-report-301481740.html.

6. Asch SE: **Studies of independence and conformity: I. A minority of one against a unanimous majority**. *Psychol Monogr: General and Applied* 1956, **70**:1–70.

7. Milgram S: *Obedience to authority*. NY: Harper; 1969.

8. * Clare DD, Levine TR: **Documenting the truth default: the low frequency of spontaneous, unprompted veracity assessments in deception detection**. *Hum Commun Res* 2019, **45**: 286–308.

Two experiments compare prompted and unprompted deception detection. The findings show that absent prompting, thoughts of honesty and deception do not come to mind.

9. Bond Jr CF, DePaulo BM: **Accuracy of deception judgments**. *Pers Soc Psychol Rev* 2006, **10**:214–234, https://doi.org/10.1207/s15327957pspr1003_2.

10. * Serota KB, Levine TR, Docan-Morgan T: *Unpacking variation in lie prevalence: prolific liars, bad lie days, or both? Communication Monographs*. 2021, https://doi.org/10.1080/03637751.2021.1985153. published online October 10.

A longitudinal study of lie prevalence is reported. Evidence of a few prolific liars is reported, and stable individual differences are parsed from day-to-day variability.

11. Nickerson RS: **Confirmation bias: a ubiquitous phenomenon in many guises**. *Rev Gen Psychol* 1998, **2**:175–220, https://doi.org/10.1037/1089-2680.2.2.175.

12. Festinger L: *A theory of cognitive dissonance*. Stanford University Press; 1957.

13. * Cialdini RB: **Influence, new and expanded: the psychology of persuasion**. *Harper Business* 2021.

This book argues for seven mechanisms underlying persuasion.

14. Arkes HR, Hackett C, Boem L: **The generality of the relation between familiarity and judged validity**. *J Behav Decis Making* 1989, **2**:81–94.

15. Effron DA, Raj M: **Misinformation and morality: encountering fake-news headlines makes them seem less unethical to publish and share**. *Psychol Sci* 2020, **31**:75–87, https://doi.org/10.1177/0956797619887896.

Five experiments document that people find it less unethical to share misinformation that that have previously repeatedly encountered.

16. * Calo R, Coward C, Spiro ES, Starbird K, West JD: **How do you solve a problem like misinformation?** *Sci Adv* 2021, **7**, eabn0481.

The authors provide a brief introduction to the distinction between misinformation and disinformation.

17. Skiba K: ***How to spot scam texts on your smartphone***. https://www.aarp.org/money/scams-fraud/info-2021/texts-smartphone.html#:~:text=In%205%20percent%20of%20overall,about%20scam%20texts%20last%20year.