Inferring truth from lies

Lauren A. Oey (loey@ucsd.edu) and Edward Vul (evul@ucsd.edu)
Department of Psychology, University of California, San Diego
9500 Gilman Dr., La Jolla, CA 92093 USA

Abstract

How much information can people gain from being lied to? We propose that people can infer the truth from false messages if two preconditions are met: (1) bigger lies are more costly, and (2) speakers have known, directional deception goals. We tested this with a marble-flipping task in which a judge tried to accurately estimate the number of sampled marbles, while a sender attempted to make the judge over- or underestimate. The sender could produce larger lies about the number of marbles drawn by physically clicking marbles along a lower or higher cost function. We found that judges took into consideration both the senders' goals and costs to correct for bias introduced by senders' lies. Our paradigm allows us to show that a large amount of the variation can be explained by people correcting others' lies based on the lies that they themselves would produce.

Keywords: deception; lie detection; truth; Theory of Mind; individual differences

Introduction

Transmitting and receiving information is a crucial goal of cooperative communication. However, people are sometimes non-cooperative communicators and endeavor to transmit false information by lying. For example, we have all experienced the tardy friend who "will be there in 5 minutes" – in this case, prior experience with the discrepancy between reality and our friend's estimate allows us to appropriately calibrate expectations. In other cases, however, we might not have any prior experience, but might still be faced with reports from a deceptive speaker. In such cases, can we still glean something about the truth from the lie that we heard?

At first glance, people may not gain much information. If the liar is just a random-number generator, or always produces the same utterance, then we learn nothing from what they have said. Furthermore, prior research suggests that people are faulty lie detectors (Bond & DePaulo, 2006), falling prey to uninformative cues about the speaker's nonverbal behavior (Vrij, Hartwig, & Granhag, 2019) or to misguided feelings of familiarity in memory (Brashier & Marsh, 2020). Cynically, we would conclude that people not only may, but should, resist updating their beliefs to messages as they cannot tease apart lies from truths, or they may believe the lie, resulting in a misbelief, which is generally maladaptive (McKay & Dennett, 2009).

Suppose, we show up at the meet up and learn that our friend does not arrive at the exact stated time. A recipient that knows a message is (logically) false learns only that our friend did not arrive in exactly 5 minutes. But this leaves open all other possibilities about what could be true about the world. Maybe our friend will arrive at the 6-minute mark. Or alternatively, perhaps people can gain much more information out of a lie, by inferring the truth. For example, we may recognize that "5 minutes" should be encoded as actually arriving in 30.

In this paper, we are concerned with how people use false statements to learn new information. In contrast, we are not concerned with how they compare statements to known facts stored in memory (e.g. Begg, Anas, & Farinacci, 1992; Fazio, Brashier, Payne, & Marsh, 2015) or how they determine if a statement is plausibly true or false (e.g. Oey, Schachner, & Vul, 2019). Additionally, we examine adversarial false messages that are intended to actively misinform, rather than cooperative pragmatic utterances, like hyperbole (Kao, Wu, Bergen, & Goodman, 2014) and polite speech (Yoon, Tessler, Goodman, & Frank, 2020), that can be literally false but are designed to be informative to the listener (Grice, 1975; Goodman & Frank, 2016). How might people spontaneously and accurately infer the truth from a lie? One way to think about how people might accomplish this is to consider the necessary preconditions.

First, we assume that lies are motivated by some goal held by the speaker. The speaker wants to induce the listener into believing something about the world that is literally false but is favorable to the speaker. For example, tardy friends may want people to believe they are more punctual than they are in reality. Therefore, tardy friends should lie by deflating their supposed arrival time. If the listener is suspicious of the speaker's goal, then estimates of the truth ought to be adjusted from the lie in the opposite direction of the speaker's goal. As a result, recipients should correct for the production bias by inferring that the tardy friend's true arrival time is inflated relative to what they reported.

Second, lies must depend on the truth. This might arise in a number of different ways. For instance, a number of cognitive models have proposed a direct relationship between speakers initially thinking about the truth and secondarily manipulating the truth to produce a lie (e.g. Walczyk, Harris, Duck, & Mulay, 2014; Debey, De Houwer, & Verschuere, 2014). On these process models, lies that are further from the truth require more cognitive effort to construct. A more general formulation is that speakers face "costs" when producing

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

lies. These costs might be due to increasing risk of detection (Oey et al., 2019), or loss of plausible deniability about their intent (Pinker, Nowak, & Lee, 2008), and moral values (Mazar, Amir, & Ariely, 2008). All these factors prevent liars from making statements completely divorced from reality. In essence, this coalesces as a cost function wherein larger lies are more costly.

Previous research has attempted to manipulate speakers' costs to lie by pushing around external factors, e.g. the presence of a third-party observer (Gneezy, Kajackaite, & Sobel, 2018; Abeler, Nosenzo, & Raymond, 2019). However, naïve listeners might not have pre-packaged expectations about how these factors scale to prevent speakers from producing larger lies. Therefore, in this study, we introduce a novel proxy for costs in the more intuitive *physical* domain, e.g. manually clicking buttons on a display. Using physical costs, we can quantitatively and in a controlled manner manipulate what costs listeners think the speaker faces. Prior studies have stuck to uniform physical costs that make it equally costly to produce a small or large lie, such as in typing lies into a text box or making a forced-choice decision (e.g. Oey et al., 2019; Gneezy, 2005). Therefore, these prior paradigms could not use physical costs for measuring how people interpret larger lies.

The current study aims to empirically test if people can infer the truth from lies, when we experimentally manipulate their knowledge about the speaker's goal and physical cost function. Participants played in a dyadic game, in which a sender draws red and blue marbles from a jar and sends a (manipulated) representation of their marbles to a judge by clicking marbles on the interface. Seeing the manipulated representation, the judge guesses how many red marbles were truly drawn. To test our proposed preconditions, we varied across participants whether the sender's goal was to make the judge *Overestimate* or *Underestimate*, and whether the sender faced a lower (`Linear`) cost or a higher (`Quadratic`) cost to produce larger lies. We then evaluated how judges' truth inferences were affected by their beliefs about the sender's goals and costs. Overall, our study informs our understanding of how people may extract information from suspected lies. If people can infer the truth from the speaker's goals and costs, that supports a broader class of theories postulating that Theory of Mind – a capacity to reason about others as rational agents responding to their own beliefs and desires – plays a critical role in people's lying and lie detecting abilities.

Experiment

Marble-Flipping Game

Participants played in a dyadic game against a computer. They were not explicitly told if their opponent was a computer or a human. Players alternated between roles as the sender and the judge. In the game, the sender drew 100 marbles[1](appearing as a jittered grid) from a virtual jar of red and blue marbles (Figure 1). The number of red and blue marbles the sender originally drew represented the truth. The sender

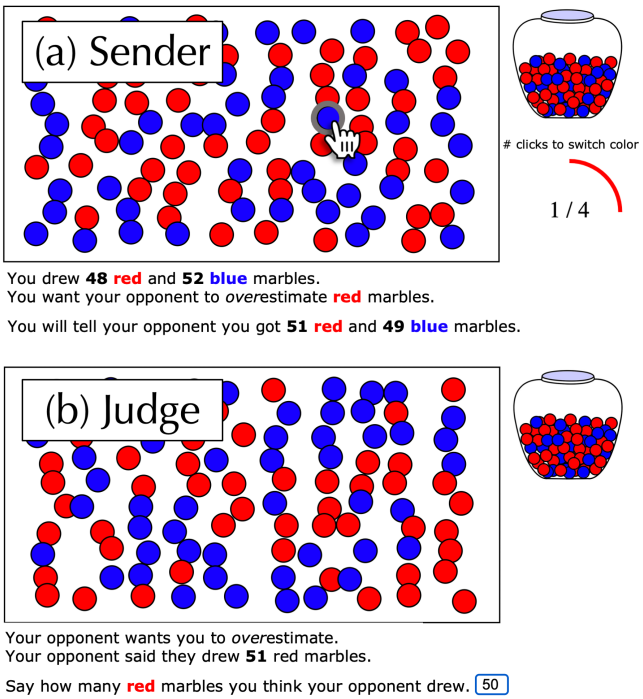

Figure 1: Game design. (a) The sender sampled marbles, and could manipulate what they showed their opponent about how many red marbles they drew by clicking marbles in the display to flip their color. The sender is told in text how many of each color marble they originally drew (e.g. "You drew 48 red...") and how many they would currently report based on their clicks (e.g. "You will tell your opponent you got 51 red..."), and a progress bar shows how many more clicks are needed to switch the next marble. (b) The judge tried to estimate how many marbles the sender truly drew from what the sender reported. In this example, the sender wants the judge to *Overestimate*, and producing larger lies follows a `Quadratic` cost function (requires additional click for each additional flip). Here, the *truth* was 48 red marbles, but the *reported* lie was 51. The judge *inferred* the truth to be 50.

then sends the judge a snapshot of how many marbles of each color they supposedly sampled. Senders had the goal to make the judge *Overestimate* or *Underestimate*; judges wanted to accurately guess what was the truth. Critically, senders could misrepresent what happened by laboriously clicking on marbles in the visual display to switch their color, before sending the report to the judge. The exact position of the marbles are randomized before the report is given to the judge, to reduce concerns about position (non)randomness as a signal to deception. The judge, seeing the display and a numerical count of red marbles, estimates the sender's original number of drawn red marbles. Participants typed their response in a text box with a valid number between 0 and 100.

[1]Participants were explicitly instructed before the task that the jar was composed of 50% red and 50% blue marbles. Not explicitly told to participants was that marbles were sampled from a beta-binomial

The sender's marble-clicking served as the manual cost to generate more extreme lies. Participants played in one of the two cost conditions. In the `Linear` condition, participants clicked each marble once to switch its color – in other words, switching n marbles requires n total clicks. In the `Quadratic` condition, participants clicked an additional time for each marble color switch, so switching the first marble required one click, the second required two clicks, the third required three clicks, etc. – switching n marbles requires $1+2+3+...+n$ total clicks. For example, switching four marbles would require $1+2+3+4$ clicks, or ten total clicks. Thus, participants in the `Quadratic` condition needed to exert more effort to produce more extreme lies. Participants were instructed before the task about the cost to switch marbles (e.g. `Quadratic`: "The more marbles you switch color, the more clicks you'll need to switch each marble.") In `Quadratic` sender trials, a circular progress bar tracked the number of clicks already completed and the number of additional clicks to switch a given marble color. In the `Linear` condition, the circular progress bar was not present.

The computer's average lying and inference behavior was held constant across cost function conditions. The computer sender lied by taking the truth and adding in the direction of their goal some sampled amount, taken from a Poisson distribution with a mean of 5. The computer receiver sampled from the same distribution but subtracted from the participant's message. Holding the computer sender's behavior constant across cost conditions ensured that any potential variation in participants' truth inference was caused by their beliefs about the sender's cost function and not by the computer sender's actual behavior.

The players' goal was to win against the other player by the largest possible point differential. Judges lost points corresponding to the absolute error of their estimate, so in Figure 1, a guess of 50 when the truth was 48 resulted in -2 points. Meanwhile, senders gained points for the judge's error in the direction of the sender's goal, so a sender who wanted the judge to *overestimate* got $+2$ points. If the judge guessed in the opposite direction (e.g. underestimated instead), the sender got 0 points, but the judge still got -2 points for their absolute error.

Participants played for two practice trials: first as the sender, then as the judge. Then, participants played for 100 test trials, switching between sender and judge roles every trial (which role was played first during the test trials is randomized). Throughout the task, participants also answered 12 attention check questions related to the trial (two in the practice trials, and ten randomly distributed in the test trials). To prevent participants from relying on learned information about their opponent's behavior, participants did not get direct feedback about their opponent's decision or the trial's outcome. To motivate participants, they received feedback about the players' cumulative points every five trials.

distribution $X \sim BetaBinomial(100,3,3)$ (95% of samples fall between 14 and 86), which allows for more variable samples relative to a standard binomial distribution (95% fall between 40 and 60). In doing so, we expected that participants would rely more on their beliefs about cost functions, rather than base-rates, to judge the truth.

Participants

Participants were recruited from the undergraduate population at University of California, San Diego to participate in an online game for course credit. Data was collected from 164 participants. Of these, 27 participants were excluded for failing to answer at least 75% of the attention check questions within a ± 5 error, five participants produced multiple responses that were out-of-bounds, and one participant had corrupted data. Additionally for participants who produce a single out-of-bounds trial, we excluded individual trials but retained the participant in our data. The remaining 131 subjects were included in our final data set. Participants were about equally distributed among the 2×2 between-subject conditions[2].

Validating preconditions in lying behavior

We confirmed the presence of our proposed preconditions in our task by examining the sender's behavior. We validated that the condition manipulations worked and senders chose lies that were driven by their assigned goals and were systematically constrained by the assigned cost function.

Senders lie consistently with their goals

We expected that senders bias their lies ($\Delta_{report-truth}$) in the same direction as their goal to either cause the judge to over- or underestimate. Using linear models with random-effects for subject and item (the true draw), we found that (as expected) senders whose goal was to overestimate on average inflated their reports relative to the truth ($\hat{\beta} = 6.14$, $t(117) = 4.65$, $p < 0.0001$), and those whose goal was to underestimate deflated their reports ($\hat{\beta} = -4.85$, $t(89) = -4.14$, $p < 0.0001$). In Figure 2 (top half of panels in white), we see that when the goal was overestimation (left panels), the majority of senders' messages deviate in the positive direction from the truth. Meanwhile, senders with the underestimation goal (right panels) generally reported numbers that were smaller than the truth (negative deviations).

Senders lie constrained by their costs

Next, we validated that the cost conditions systematically influenced how senders lied. If the amount of effort senders committed to trials was consistent between the conditions, then we would expect that senders produce greater bias in their lies when subjected to lower costs in the `Linear` condition. Indeed we found that the linear cost senders introduced more bias into their reports relative to the quadratic cost senders ($\hat{\beta} = 5.11$, $t(129) = 3.47$, $p < 0.001$), aggregating over goals. In Figure 2, we see that senders with a `Linear` cost (top panels) on average swapped about 7.4 marbles in the expected direction, while in the `Quadratic` cost (bottom

[2]Data and code for the experiment and analysis are available at `https://github.com/la-oey/WhatIsReality`

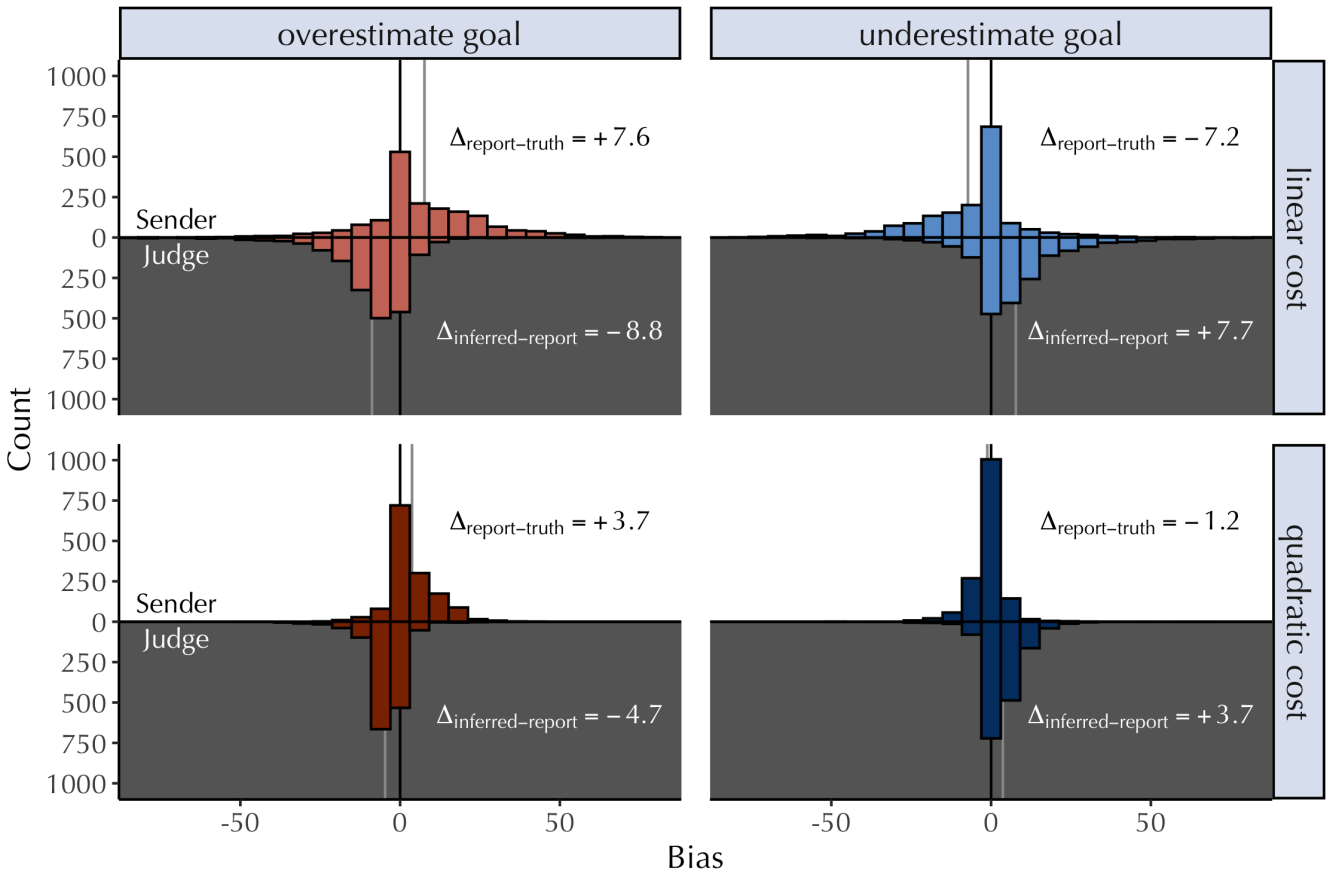

Figure 2: Distribution of participant senders' biasing and judges' bias-correcting behavior across each condition (panel columns are goals and rows are cost functions). The top half of each panel (in white) shows how much senders manipulate their report relative to the truth ($\Delta_{report-truth}$). The bottom half (in gray) shows how much judges adjust their inferred truth relative to the sender's report ($\Delta_{inferred-report}$). The direction and distance of the gray line (mean bias) relative to the black line at $bias = 0$ indicates if participants generally inflate (positive bias) or deflate (negative bias) their response and how large the difference is.

panels) condition they only swapped about 2.5 marbles on average. Additionally, there is more variance in bias produced under the `Linear` cost function.

Do people estimate the truth by considering their beliefs about others' goals and costs?

Given that participants produced lies that were guided by their own goals and costs, did they also infer the truth by applying their beliefs about the senders' goals and costs?

Judges correct for bias considering goals

If judges applied their beliefs about senders' goals, we would expect them to make bias corrections ($\Delta_{inferred-report}$) that were in the opposite direction of the senders' goal. If the sender wanted the judge to overestimate, then the judge should expect the sender to positively bias their report by adding more red marbles. A judge who expects positive bias in the report should correct for the bias in the negative direction by guessing that the true number of marbles drawn was fewer than what was reported.

As predicted, we found that participants in the *Overestimate* condition bias-corrected in the negative direction by guessing smaller numbers ($\hat{\beta} = -6.96$, $t(100) = -8.17$, $p < 0.0001$; Figure 2 left panels). Vice versa, participants in the *Underestimate* condition bias-corrected in the positive direction by guessing larger numbers ($\hat{\beta} = 5.75$, $t(89) = 5.12$, $p < 0.0001$; Figure 2 right panels). These results show that the direction of people's truth inferences are informed by their beliefs about speakers' goals.

Judges correct for bias considering cost function

Judges that apply their beliefs about senders' cost functions should expect senders with lower cost functions to produce more extreme lies. Therefore, they should make larger magnitude bias corrections. Indeed, judges who believed the sender had a `Linear` cost debiased their guess more compared to the `Quadratic` cost ($\hat{\beta} = -4.24$, $t(129) = -3.51$, $p < 0.001$), aggregating over goals. Figure 2 shows that the bias corrections' absolute distance to the intercept is larger for the `Linear` cost (top panels), compared to the `Quadratic` cost (bottom panels).

Human judge correction relates to sender bias

What mechanisms might be driving how people decide to interpret lies? There has been a recent push in the literature toward using recursive pragmatic frameworks to understand communication when speakers and listeners have misaligned goals (e.g. Ransom, Voorspoels, Navarro, & Perfors, 2019; Oey et al., 2019). A speaker reasons about how a listener might interpret an utterance influenced by the listener's beliefs and goals, and the listener in turns reasons about how a speaker ought to produce an utterance under *the speaker's* beliefs and goals. Unfortunately, people are not telepathic, so they need to conjure a model of how their opponent ought to think and behave, and decide their own behavior under that normative model.

A naïve rational heuristic might be to anchor the model of their opponent to the easily accessible model of oneself. In the marble-flipping task, a participant might just assume that the computer sender biases their lies to the same extent that she does when she is the sender, and so as the judge, she should bias-correct to the same magnitude. In other words, pragmatic frameworks suggest a systematic relationship of individual differences binding people's sender and judgment behavior (for a similar phenomenon in persuasion, see Barnett, Hawkins, & Griffiths, 2021). Here, we evaluate the relationship between human sender biasing and human judge bias-correcting, both in aggregate and in individuals.

In aggregate

If people assume their opponent behaves like themselves, we would expect a similar (but mirrored) distribution between how human judges correct for bias and how senders bias their reports. This is what we find in Figure 2 – the patterns of the sender's biasing (top half) and the judge's bias correcting (bottom half) appear to be rotationally symmetrical. Visually this highlights a similar skewed shape and spread in the bias data.

We can also compare the bias means. Figure 3 plots the means for each condition (as rhombuses) in a 2D space, with the x-axis as senders' bias and the y-axis as judges' bias correction. If judges flip the sign of their bias, then we would expect that the *Overestimate* bias means would be located in the top left quadrant and *Underestimate* means would be in the bottom right quadrant. Additionally, we would expect the `Quadratic` means to be shifted toward the origin, relative to the `Linear` means. These qualitative patterns are what we find. Quantitatively, examining the relationship between human senders' bias and human judges' bias by focusing on the condition means, we find a very strong negative correlation of $r = -0.98$ ($t(2) = -7.77$, $p < 0.02$). Between conditions, when people produce larger bias in their reports, they correct more in the opposite direction as a judge.

In individuals

Focusing on individual differences would provide stronger evidence that people anchor their truth judgments to their own report behavior as a sender. We examined the relationship between individual senders' bias and judges' bias correction (scatter points in Figure 3). Correcting for aggregated means, we found a significant negative correlation of $r = -0.57$ ($t(129) = -7.79$, $p < 0.0001$). An individual sender's mean bias determined 32% of the variation in their mean bias correction as a judge. These results suggest that a large contributor to what people are doing is modeling their opponent based on themselves.

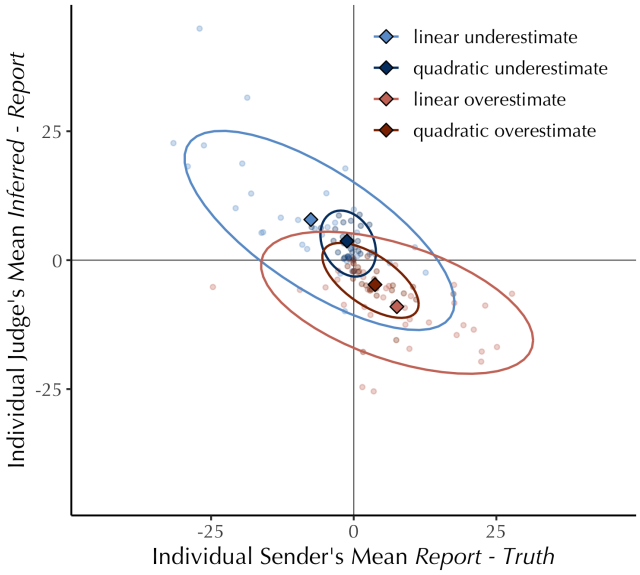

Figure 3: Individual differences in sender and judge behavior. Sender behavior (x-axis) is summarized as the mean introduced bias ($Report - Truth$), and receiver (y-axis) is the mean bias correction ($Inferred - Report$). Scatter points represent individual participants, with rhombuses showing the mean for each condition, and ellipses showing spread. Individuals' judge behavior negatively correlates with their sender behavior, so people who produce larger lies assume their opponent also produces larger lies. The means of the *Underestimate* (and *Overestimate*) goal conditions are in the top left (and bottom right) quadrant, showing that people both biased and bias-corrected in the predicted direction. There is less spread and means are shifted toward the origin in `Quadratic` (relative to `Linear`) cost conditions.

Discussion

In sum, this paper explored how people may infer the truth when lied to. At first glance, this seems like such an ability would require omniscience. However, we proposed, and experimentally showed, that when the necessary preconditions are fulfilled – namely that (1) lies are directed by speakers' goals, and (2) lies are constrained by cost function – people may infer the truth from lies. We validated that participants' lying behavior reflected their goals and cost functions. Critically, when inferring the truth, judges showed sensitivity to the senders' goals and cost functions by tuning their infer-

ences about the underlying reality accordingly.

In this paper, we focus on one aspect of deception – that a speaker's goal is for the receiver to mis-estimate in a certain direction (e.g. I am putting in maximal effort to be punctual), while a receiver's goal is to infer the truth (e.g. my friend will actually be X minutes late). Built into this study's task is the high suspicion that messages are lies. In fact, participants are explicitly instructed about the speakers' goal to induce the receiver to mis-estimate. Additionally, ground truth follows a high variance distribution so that participants rely more on beliefs about costs rather than base-rates (e.g. the report is suspiciously far from 50). Here we ignore another typical goal in deception: avoiding suspicion to make the receiver believe that the speaker is telling the truth (Oey et al., 2019). Future work should explore how both goals combine to form the kinds of lies people produce in the real world.

Another crucial underlying assumption in this paper is that some lies are larger than others. What constitutes a large lie in mental representation? In this study, we measure larger lies in terms of bias: for the sender, it is the scalar difference between what they drew versus what they reported. However, for many naturalistic settings, the magnitude of a lie cannot be measured on a number line. For example, if one reports their income as $\$1,000,000$ instead of $\$100,000$, a listener could plausibly excuse the lie as a surplus "0" rather than a lie of $+\$900,000$. How do people generate hypotheses about why a false statement might deviate from reality?

The design of our study allowed us to evaluated a prediction from pragmatic accounts of adversarial communication – that people might infer the truth by anchoring their model of the opponent to their own lie-telling behavior. The account predicts systematic individual differences between how people tell lies and how they infer the truth, and (teasing out the variation explained by conditions) we found that it explained a large portion (32%) of the remaining variation. What else might explain the remainder of the variation? One asocial explanation is that people who strongly distrust others may over rely on their prior beliefs about what could be true and entirely ignore false messages. Another more Theory of Mind intensive explanation is that people may expect others to have systematically deviating behavior from themselves and calibrate their behavior to agents' unique beliefs (Oey & Vul, 2021). To better pinpoint Theory of Mind as a critical mechanism for inferring the truth, future work should examine how people vary their inferences when they believe the speakers has alternative beliefs.

Acknowledgments

This material is based upon work supported by the NSF Graduate Research Fellowship under Grant No. DGE-1650112 to LAO. We thank Chaz Firestone whose anecdotal tweet from April 29, 2021 (about his doctor's visit) inspired our research question. We also thank Eric Xiao for programming some functions for the task.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153.

Barnett, S. A., Hawkins, R. D., & Griffiths, T. L. (2021). A pragmatic account of the weak evidence effect. *PsyArXiv*.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*, 499–515.

Debey, E., De Houwer, J., & Verschuere, B. (2014). Lying relies on the truth. *Cognition*, *132*(3), 324–334.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384–394.

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, *108*(2), 419–453.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818-829.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics Vol. 3: Speech Acts* (pp. 64–75). New York: Academic Press.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, *32*(6), 493–510.

Oey, L. A., Schachner, A., & Vul, E. (2019). Designing good deception: Recursive theory of mind in lying and lie detection. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 897–903). Austin, TX: Cognitive Science Society.

Oey, L. A., & Vul, E. (2021). Lies are crafted to the audience. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society* (pp. 791–797). Vienna, Austria: Cognitive Science Society.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, *105*(3), 833–838.

Ransom, K., Voorspoels, W., Navarro, D. J., & Perfors, A. (2019). Where the truth lies: How sampling implications drive deception without lying. *PsyArXiv*.

Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: Nonverbal communication and deception. *Annual Review of Psychology*, *70*, 295–317.

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22–36.

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, *4*, 71–87.