Routledge
Taylor & Francis Group

Communication Monographs

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rcmm20

Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both?

Kim B. Serota, Timothy R. Levine & Tony Docan-Morgan

To cite this article: Kim B. Serota, Timothy R. Levine & Tony Docan-Morgan (2022) Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both?, Communication Monographs, 89:3, 307-331, DOI: 10.1080/03637751.2021.1985153

To link to this article: https://doi.org/10.1080/03637751.2021.1985153

COMMUNICATION MONOGRAPHS
2022, VOL. 89, NO. 3, 307–331
https://doi.org/10.1080/03637751.2021.1985153

Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both?

Kim B. Serota [a], Timothy R. Levine[b] and Tony Docan-Morgan [c]

[a]School of Business Administration, Oakland University, Rochester, MI, USA; [b]Communication Studies, The University of Alabama at Birmingham, Birmingham, AL, USA; [c]Department of Communication Studies, University of Wisconsin-La Crosse, La Crosse, WI, USA

ABSTRACT

Testing truth-default theory, individual-level variation in lie frequency was parsed from within-individual day-to-day variation (good/bad lie days) by examining 116,366 lies told by 632 participants over 91 days. As predicted and consistent with prior findings, the distribution was positively skewed. Most participants lied infrequently and most lies were told by a few prolific liars. Approximately three-quarters of participants were consistently low-frequency liars. Across participants, lying comprised 7% of total communication and almost 90% of all lies were little white lies. About 58% of the variance was explained by stable individual differences with approximately 42% of the variance attributable to within-person day-to-day variability. The data were consistent with both the existence of a few prolific liars and good/bad lie days.

ARTICLE HISTORY
Received 1 March 2021
Accepted 8 September 2021

KEYWORDS
Deception; lies; lying over time; prevalence; prolific liars; truth-default theory; TDT

Research has found that people lie, on average, about once or twice per day (DePaulo et al., 1996; Serota et al., 2010). It is increasingly apparent, however, that the average number of lies per day reported in the literature does not reflect the behavior of the majority of people. The distribution of lying is highly skewed (Serota et al., 2010). Most people report telling few or no lies on a given day and most lies are told by only a few prolific liars. Over the past decade, the skewed distribution of lie prevalence has emerged as an exceptionally robust phenomenon (Levine, 2020). Nevertheless, the skewed lie frequency distribution may still be partially misunderstood because prior findings come from cross-sectional data or data collected over a relatively short time frame.

Self-reports of the frequency of lying are often interpreted as reflecting stable individual differences between people. It is tacitly presumed that the more-honest-than-average majority and the prolific liars who are out on the long right-hand tail of the distribution are different sorts of people. That is, the current understanding is that prolific liars are distinct and potentially identifiable people with particular characteristics that manifest through consistently telling an unusually large number of lies relative to the majority of people. Evidence of this tacit presumption is reflected in efforts to link prolific lying

CONTACT Timothy R. Levine levinet111@gmail.com
Supplemental data for this article can be accessed https://doi.org/10.1080/03637751.2021.1985153.

with individual difference variables such as specific personality traits or demographic characteristics (e.g., Michels et al., 2020). This presumption may not be warranted and may reflect the actor-observer asymmetry in attributions in which the negative behaviors observed in others (e.g., frequent lying) are ascribed to dispositional rather than situational causes (Malle, 2006).

Typically, lie prevalence has been measured at a single point in time. With cross-sectional data, it is unclear whether or not the people who lie prolifically on any given day do so on most days. Perhaps those who report a large number of lies on the day of the survey are just having an unusually bad lie day. Lying prolifically on the day of the survey may not reflect their usual lie behavior or their dispositional tendencies. Conversely, people who tell lots of lies on most days may have days on which they tell few lies (a "good lie day"). In sum, being a prolific liar and lying prolifically on a given day are distinct concepts that are conflated in cross-sectional data.

It is very likely that some people consistently lie more than others over time. It is also almost certain that individuals vary from day-to-day in the number of lies they tell. Disentangling the distribution of lie prevalence among different people from the distribution of within-person lying over time requires tracking individuals' lie behavior over time. The current study sought to resolve this ambiguity by tracking individuals daily over multiple months. Based on truth-default theory (TDT; Levine, 2014, 2020), we predicted that both the few prolific liars and the bad lie days hypotheses would explain meaningful variation in lie prevalence.

Lie prevalence research

DePaulo et al. (1996) produced the first influential (albeit small sample) measurement of lie prevalence finding that, on average, Americans lie once or twice a day. This finding has become widely cited (see Docan-Morgan, 2019), often with the interpretation hinging on the implicit presumption that the average reflected the typical lie behavior of most people.

Serota et al. (2010) tested the 1–2 lies/day claim by surveying 1000 American adults. Although the average was replicated ($M = 1.65$ lies/day), the study found that lying is not normally distributed across subjects; 59.9% told no lies and half of all reported lies were told by just 5.3% of subjects. That is, the average was indeed between one and two, but the majority of the adults sampled told zero lies on the day of the survey. The average did not reflect typical behavior and lying was not an everyday behavior for most people. The average was distorted by "outliars," extreme scores produced by a relatively few prolific liars.

Serota et al. (2010) subsequently examined data from several prior studies and found a similar skewed pattern in which most subjects told few lies and a few subjects told a substantial portion of lies. Student written diaries (see the original DePaulo et al., 1996 study), electronic diaries (George & Robb, 2008), and experimental research (Feldman et al., 2002) all showed similar skewed distributions. Serota et al. also replicated the shape of the distribution from the national adult study with a student sample. Most students told 0–2 lies per day and 5.8% of the sample accounted for nearly a quarter of all lies told. Notably, the mean for students ($M = 2.34$) was higher than the mean for adults ($M = 1.65$) showing, as did DePaulo et al.'s (1996) original data, that college students lie more

than adults. In fact, age has emerged as a robust predictor of lie frequency. Lying increases with age through childhood, peaks during the teens, and then declines gradually with age (Debey et al., 2015; Serota et al., 2010). Nevertheless, the skewed distribution is robust across age groups (Debey et al., 2015; Serota et al., 2010).

The few prolific liars pattern has subsequently been replicated internationally and across communication modalities. Halevy et al. (2013), using a student sample and adding behavioral validation of self-report measurement, and Debey et al. (2015), with an age cross-section ranging from children to seniors, observed similar patterns in the Netherlands. Serota and Levine (2015) replicated the lie distribution finding with 2980 subjects in the U.K., including replications in the culturally variant England, Wales, Scotland, and Northern Ireland sub-samples. Stockman (2017) observed the long-tail distribution in Belgium, and a re-examination of data reported by Zvi and Elaad (2018) found the same distribution among Israeli subjects. Most recently, the skewed lie distribution has been replicated in South Korea (Park et al., 2021) and Japan (Diaku et al., 2021). In studies of mediated communication, the skewed distribution of lie frequency has been reported in both general text messaging (Smith et al., 2014) and mobile dating situations (Markowitz & Hancock, 2018).

It is worth noting that the long-tail distribution has been obtained regardless of how the lie question is asked. Serota et al. (2010) used an elaborated question with multiple categories (i.e., who was lied to by medium). The London Science Museum survey reported by Serota and Levine (2015) asked respondents to report the number of lies told on a typical day in two broad categories: big lies and little white lies. Many studies, especially those for which lie frequency is just one of many deception measures, ask for a single number indicating the total number of lies told in the previous 24 h (e.g., Debey et al., 2015). The text message study by Smith et al. (2014) looked at the rate of lies, with lies calculated as a share of total messages.

Although the various study means are indicative of how much lying occurs and can be used to compare subsets such as age groups or other characteristics within studies, across studies the means are most conspicuous for the narrow range in which they have been observed. The consistency of the distributions when observed across samples, methodologies, and the questions themselves is more relevant to the hypotheses of this study.

Truth-default theory

TDT (Levine, 2014, 2020) is a modular theory of human-to-human honest and deceptive communication. The scope of TDT includes both the sender and receiver sides of communication. Although the current research focus is on the sender side, the broader theory is briefly summarized to provide theoretical context.

According to TDT, believing the content of others' messages is necessary for efficient communication. By default, people believe other people. The conscious recognition that others' messages might be false or deceptive does not come to mind unless suspicion is actively triggered. Consistent with the truth-default, recent studies have reported few deception-related thoughts when viewing false or deceptive content in the absence of a strong trigger (e.g., Levine et al., 2020). The truth-default is argued to be overwhelmingly advantageous for individuals and for the species because it facilitates effective communication. Defaulting to the truth, however, involves a trade-off as vulnerability to deception

is increased in the short term. Critical for the current discussion, the gain-harm implications of the truth-default trade-off hinge on both the frequency of deception and the severity of the lies. The less frequent and less consequential the lying, the more value is added by defaulting to the truth.

TDT lists a variety of specific triggers that prompt suspicion and deception judgments. TDT further identifies types of information that are more or less useful in lie detection. Unlike many theories of deception, TDT holds that reliance on deception cues hinders deception detection accuracy. Most deception is detected after the fact based on communication content, external evidence, or honest confessions. In further contrast to conventional wisdom, TDT holds that truth-bias actually improves the truth-lie hit rate (cumulative raw truth-lie judgment accuracy) because most communication in our everyday lives is honest. Thus, in this way too, the logic of TDT hinges on the prevalence of deception in human communication.

TDT has two modules on the sender side of interaction. Together, the sender-side modules propose that most people lie infrequently and that people lie only as-needed to accomplish communication goals that cannot otherwise be achieved through honesty. The few prolific liars module contains three interrelated conjectures: overall lying is infrequent relative to honest communication (i.e., the truth-lie base-rate in normal interaction tilts heavily to honesty), the frequency of lying is not normally distributed across the population, and most lies are told by a few prolific liars.

The deception motives module proposes that a different sort of truth-default applies to message production, namely that people are honest unless they have a reason to deceive. That is, honesty is the default mode of communication. Communication is often viewed as a goal-directed activity (Palomares, 2014). In TDT, people communicate honesty unless the truth of a situation thwarts the goal. Deception is a tactical means to overcome an obstacle to goal attainment. For example, Levine et al. (2010) reported that across several situations, when the truth aligns with communication goals, people are invariably honest. Deception is limited to situations where the truth is problematic.

It follows from the two sender-side modules that lie prevalence has both individual difference and situational antecedents. The few prolific liars module specifies how people are distributed in terms of lie frequency. The deception motive module holds that lies are told under predictable and understandable circumstances. The situations that motivate deception need not be uniform from day-to-day for any given individual. Together, these propositions suggest that there are indeed prolific liars, and prolific liars are unusual in comparison to the much more honest majority. Individuals, however, are also likely to show day-to-day variability because each individual lie is situationally determined. Although prior research has supported each prediction individually, it has yet to track individual lie behavior over an extended period of time, and thereby assess individual and within-individual variability in tandem.

A proof-of-concept case study: Fact-checking Donald J. Trump

Prior data on lie prevalence come predominantly from cross-sectional surveys or diary studies conducted over a relatively short time frame (e.g., one week, DePaulo et al., 1996; one day, Serota et al., 2010). Yet longer-term data is available for one particular individual: The 45th President of the United States of America, Donald J. Trump.

From his inauguration to the end of his presidency, *The Washington Post* Fact Checker recorded the number of public false and misleading statements of Donald Trump on a daily basis. These data provide a unique peek at an unusually prolific prevaricator, thereby inspiring the current effort to parse individual differences from day-to-day variability.[1]

The distribution of Trump's false and misleading statements over time displays large day-to-day variability and, as Figure 1A shows, an average that changes over time (Serota, 2019). The 30-day moving average during the first part of his term shows a rising trend from Inauguration Day, 2016 to the end of October 2017, but the lie activity of one day does not predict subsequent activity for any of the first 16 autocorrelation intervals (Ljung–Box test for lack of fit, $p < .001$ for all lag levels). However, when the frequency (number of days) for each number of lies is plotted (see Figure 1B), the pattern resembles those observed in cross-sectional studies of lies told on a given day (Serota et al., 2010; Serota & Levine, 2015). Trump's daily frequency, from one to 125 lies, fits a power function ($y = 174.23x^{-1.185}$, $R^2 = 0.88$) similar to those observed in numerous cross-sectional studies. If the Trump pattern is representative, it suggests that individual behavior over time is similar to a single day's behavior distributed over many individuals.

The findings from the fact-checking data document that individual behavior varies considerably from day-to-day and most likely does so in response to situational influences. Observed over time, Trump appears to lie considerably more than the majority of Americans. His mean of 9.9 lies per day during the time frame we examined puts him above the 99th percentile of Americans in the Serota et al. (2010) study. However, there are two observations critical to the current discussion. First, Trump's mean lies per day depends on the time frame sampled. For example, his average during the first 30 days sampled was 4.6 lies/day compared to 36.8 lies/day in the final thirty days sampled. Second, when we look at a one-day snapshot of Trump's lies, he might not be revealed as an unusually frequent liar. Of the 649 consecutive days examined, Trump lied 0, 1, or 2 times (i.e., values within the normal range) on 38.2% of those days. In contrast, he told five to fifteen lies (his average, +/−5 lies) on 27.1% of the days. In short, depending on which day was sampled, his lie frequency might be zero

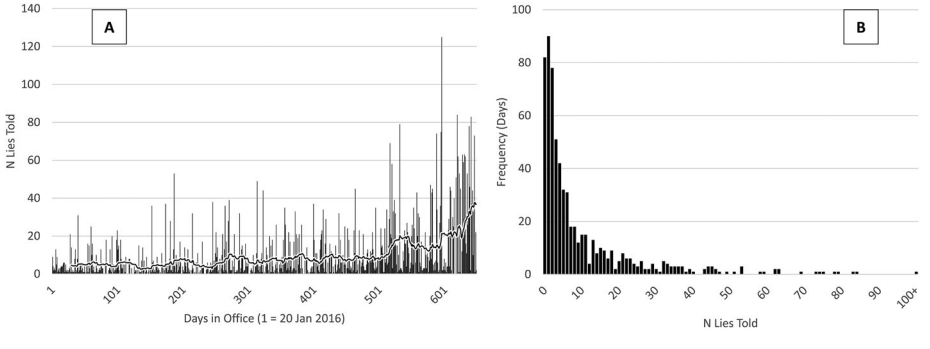

Figure 1. Number of fact checked false statements told by Donald Trump over time. Notes: (A) Vertical bars indicate the number of lies told by Donald Trump each day from 20 January 2016 to 30 October 2017 ($n = 649$ days; $M = 9.9$, $SD = 14.73$); data from *The Washington Post* Fact Checker. The solid line is the 30-day moving average. (B) The same data, grouped and arrayed from 0 to 100+ falsehoods told (*Skew* = 2.90).

or it could be 125. We would draw very different conclusions about his lying behavior depending on the particular day that was sampled, and his tendencies are only recognizable when viewed over time. This insight motived the current investigation into lying over time. Could the same day-to-day variability be true of other people?

Predictions

The basic thesis of the current research is that the skewed distribution of lie prevalence observed in cross-sectional data is a joint function of individual differences in lie behavior that persist over time and within-person temporal variation. This thesis follows from integrating the few prolific liars and the deception motive modules of TDT. The need to integrate the two modules is apparent from the Trump fact-checking data. The Trump data suggest that investigating the modules independently might produce an incomplete understanding of lie prevalence because each module holds implications for the other. It follows from the temporal, day-to-day situational variation anticipated by the deception motive module that cross-sectional data provides an insufficient test of the few prolific liars module because lie behavior on a given day need not reflect behavior typical of the individual. Conversely, the individual differences specified in the few prolific liars module suggest that individuals are differentially likely to face various deception motives at different rates of occurrence. For example, individuals higher in narcissism may be more likely to encounter situations that thwart their impression management goals requiring them to deceptively inflate their self-image. People who enact more antisocial behaviors need to lie more to cover their transgressions.

When individual lie behavior is examined over time, we expect evidence for both (1) a few prolific liars as a distinct and identifiable set of individuals, and (2) for "bad lie days" in which individuals lie more (or less) than what is typical of their own personal baseline. That is, we expect a pattern of results consistent with both the few prolific liars module and the deception motive module of TDT:

> H_1: *(the few prolific liars hypotheses)*: (H_{1a}) In aggregate, lie prevalence data is positively skewed with most people lying infrequently and most lies being told by a relatively few prolific liars. (H_{1b}) On any given survey day a similar pattern will be observed. (H_{1c}) This variation originates at least partially at the individual level (individual differences).

> H_2: *(the bad and good lie days hypotheses)*: (H_{2a}) People vary from day to day in the number of lies they tell. Consequently, (H_{2b}) cross-sectional data can misclassify prolific liars.

Additionally, we offer four research questions. H_1 drew a basic distinction between the majority normatively honest communicators and the numerically smaller group of prolific liars. Perhaps longitudinal data allow for the further parsing of individuals into groups or clusters based on similarities and dissimilarities in lie behavior:

> RQ_1: Can people be grouped based on similarities and dissimilarities in lie behavior over time?

Second, in the Trump case study, the distribution of Trump's lies revealed a positive skew similar to the distribution observed across individuals. Of course, generalizing from $N = 1$ data is ill-advised, but this observation raises the question of the extent to which it may or may not be idiosyncratic:

RQ_2: Will the skewed distribution of lie frequency predicted *across* individuals also extend to *within* individuals?

It is sometimes presumed that "everyone lies." In the original Serota et al. (2010) national survey (Study 1), nearly 60% of participants reported telling no lies in the past 24 h. In a follow-up survey with students (Study 3), 28.9% reported telling no lies in the past 24 h but 92.4% reported lying at least once in the past week. In the DePaulo et al. (1996) diary study, one student participant (1.4%) and six adult participants (8.6%) reported telling no lies over one week. Thus, prior studies with both student and adult samples have reported data in which some participants denied lying. When we collect data over a longer time frame, will everyone lie at least once, or might there be participants who go months without lying?

RQ_3: Will everyone lie?

Long-term longitudinal studies can be informative and may facilitate testing hypotheses such as those proposed here. However, as a practical matter, extensive repeated measures are unwieldy to execute. In order to measure lie frequency in conjunction with other behavioral and attitudinal measures, most surveys and experiments conducted to date have been cross-sectional in nature. One alternative to using a single day measure is to obtain a self-estimate of average lies per day. Can we estimate the lie rate at a single point in time?

RQ_4: Is an individual's estimate of their average lie behavior a good indicator of an individual's normative behavior?

Method

Participants

The core sample for this study consisted of 632 undergraduate students enrolled in communication courses at a medium-sized university in the Midwestern United States. Students were enrolled in either a required general-education communication course or in courses for communication majors or minors. Participants reported their sex as male ($n = 158$; 25.0%), female ($n = 472$; 74.7%), or not disclosed ($n = 2$; 0.3%). Although the sample was disproportionally female, we test-weighted key variables and determined weighting sex to either the sampling frame or the U.S. population would have only a small effect and not affect our conclusions. Ages ranged from 18 years old to the category 31–40 years old ($M = 19.1$, $SD = 1.59$; *Median* and *Mode* = 18 years old).

The large-scale longitudinal nature of the study prescribed the use of students who could be tracked over an extended time period. The focus of this study was variation rather than mean levels of lying. Thus, differences between college age and adult samples, which had been established previously and affect mean differences rather than the shape of distributions (Debey et al., 2015; DePaulo et al., 1996; Serota et al., 2010), are less critical. We accede that college students tell more lies than adults.

Procedure and measures

After receiving institutional review board approval, students enrolled in communication courses during the Fall 2019 semester were invited to participate in the

current study. During the first week of the semester, 25 instructors across 84 sections of courses invited their students to participate in the current study. The pool was approximately 1400 students. Announcements were made orally in class, via email, and posted in course learning platforms. Students were informed that the purpose of the study was to examine how often people lie and would include completing a series of daily online questionnaires via smartphone, computer, or tablet. As an incentive, students were informed that if they participated in the study for the duration of the semester, they would receive extra credit or course credit as partial fulfillment of a requirement for their course. To avoid attrition, students were also entered into multiple raffles to win $25 gift cards during the last six weeks of data collection. Students who chose not to participate were given the option to complete an alternative assignment. The study was administered through a sign-up survey and four phases of questionnaires.

Sign-up survey

Students who wanted to learn more about the study and/or sign up to participate accessed a link to Qualtrics, a web-based survey software program. This sign-up survey reiterated the purpose of the study, stated general expectations and included informed consent information. At the end of the survey, students were able to opt in or out of the study. Students also reported their age and contact information for future study communications and participant tracking. After eliminating duplicate, incomplete, opt-out, and non-qualified (participants under 18 years old) surveys, 844 participants were sent survey links for Phases 1 and 2. After the second wave of opt-out/non-response eliminations, 819 participants were enrolled in Phase 3, the daily lie survey. The final sample consisted of 632 participants who completed a minimum of 85 of the 91 daily surveys and Phase 4.

Phase 1

Students who opted in to participate were emailed a link to complete the first phase of the study. Phase 1 included demographic information and questions about perceptions of lying.

Phase 2

Upon completion of Phase 1, participants were emailed a link to complete Phase 2 and reminded about the confidentiality of their responses. As part of the larger dataset for this project, students completed numerous measures pertaining to deception, language use, personality, and values. Participants were asked how many times they lied in the past 24 h (using the Serota et al., 2010 question format) and asked to estimate how many times they lied during a typical day. Finally, participants were introduced to the third phase of the study.

Phase 3

During this continuous panel phase, participants completed a measure of lie behavior for 91 consecutive days. On a daily basis, participants completed one of five randomly assigned measures of daily lie behavior. To encourage participation, the majority of data was obtained using a single-number recall question. As discussed in the

introduction, other forms can be used to obtain additional information about the nature of the lies told. Over time, an individual participant may have received 1, 2, 3, 4, or 5 versions; the alternative versions are explained below. During Phase 3, each participant received a daily email with a link to the survey and a link to opt out. On the first day, all participants read the following prompt (on subsequent days, the prompt was not shown but was available by clicking a link):

> Most people think a lie occurs any time you intentionally try to mislead someone. Some lies are big while others are small; some are completely false statements and others are truths with a few essential details made up or left out. Some lies are obvious, and some are very subtle. Some lies are told for a good reason. Some lies are selfish, other lies protect others. We are interested in all these different types of lies. To help us understand lying, we are asking many people to tell us how often they lie.
>
> Think about where you were and what you were doing during the past 24 h, from this time yesterday until right now. There are many kinds of people you might have lied to … family, friends, classmates or work associates, acquaintances, and total strangers. You might have talked to them face-to-face or some other way such as in writing or by phone or over the internet. Think about all of these situations as you consider how many times you lied. If you have not told any lies, we want to know that, too.

The alternative question wordings included a one-item response and four, more detailed questions developed from prior research. Initially, all respondents received the one-item measure and the additional versions were phased in over time. The phased start allowed each version to be introduced separately as part of the daily survey emails. The version a respondent received on any given day was determined by random assignment. The one-item question was given priority weighting in the random assignment scheme.

One-item daily lie measure (88.4% of 57,310 survey days). On most days, participants were sent and completed a one-item lie question: "In the past 24 h, how many times have you lied? Write in one number for your total lies. If you told no lies, write in "0"." On average, this measure took participants less than one minute to complete.

10-item daily lie measure (5.9%). Participants were randomly assigned to answer an alternate daily lie measure. One of these measures asked participants to report how many times they lied in the past 24 h for each of the following 10 contexts (adapted from Serota et al., 2010): family members – face-to-face (F2F), family members – written or by phone or Internet (not F2F), friends or other people you know socially (F2F or not F2F), classmates/teachers and people you work with or know as business contacts (F2F or not F2F), people you don't know but might see occasionally (such as a store clerk) (F2F or not F2F), and total strangers (F2F or not F2F).

White lies versus big lies (2.2%). The second alternate daily lie measure asked participants to report their white lies and big lies (adapted from Serota & Levine, 2015):

> In the past 24 h, how many times have you lied? Some lies are little white lies; others are big lies (what a person considers to be a big lie is a personal judgment, so you will have to decide for yourself which of your lies are little or big). In the boxes below, please write in the

> number of times you told each type of lie. Only fill in boxes for the kinds of lies you told. The total will be calculated automatically.

Measure of daily lies and percent of communication (1.9%). The third alternate version combined the one-item daily lie measure with a question regarding the reported number of lies as a proportion of participants' total daily communication. Participants who reported one or more lies on that day were presented with the following question:

> Thinking about all of *your* communication in the past 24 h, the lie(s) you reported above was (were) what percent of your total communication to others? (Some people think about communication in terms of time or messages or interactions; answer the question using the way you typically think about communicating.)

Participants selected a percentage using a slider scale ranging from 0% to100%. Participants reporting no lies in the past 24 h were not asked the follow-up question.

Measure of lie motives (1.6%). The final alternate daily lie measure asked the following:

> In the past 24 h, how many times have you lied for the reasons listed below? In each of the boxes below, please write in the number of times you have lied for this reason. Only fill in boxes for the reasons you lied. The total will be calculated automatically. If you feel you told a specific lie for more than one reason, do not count the same lie more than once. Instead, count it for the main reason or most important reason you told that lie.

Motive items were adapted from Levine et al. (2016) and included the following: to avoid others, to protect yourself, to protect another person, for your own benefit or gain, for the benefit of another person, to impress or appear more favorable, for humor (as a joke or prank), to hurt another person, for no reason (it's just something you do), and for a reason not listed (write in).

Of the 819 students who agreed to participate initially, those who opted out or completed fewer than 85 days of lie reporting were removed from the database. Total lies were calculated by aggregating the responses to the alternate questions to obtain a single number for each response (11.6% of total survey days) and combined with the one-item question responses. Statistics were calculated for the distribution of each participant's daily total reports over the 85–91 successive measures.

Phase 4

At the end of the semester, participants completed several individual differences measures and responded to additional questions regarding expectations about lying and their experience with the survey. Participants who did not complete Phase 4 were deleted from the sample; 632 participants completed all four phases of the study.

Results

Distribution of lies for total survey days

The key measurement for this study was repeated daily self-reporting of total lies. Participants reported lying in the past 24 h each day for three months. Reports using one of the multi-category questions were summed to obtain the total. Statistics were calculated for

the overall distribution (testing H_{1a}), the distribution across subjects for each of the 91 days (testing H_{1b}), and the distribution of each participant's reports over the 91 successive day measures (testing H_{2a}). Given the potential for missing reports over three months, an a priori decision was made at the start of the daily surveys to allow up to one week of non-responses. Participants with fewer than 85 days of completed reporting were removed from the database. In all, 632 participants over 91 days provided a sample of 57,310 survey days, the aggregate daily reports of lies (this excludes 202 non-responses or 0.35% of the possible participant-by-day reports). Across the total survey days, we observed a grand mean of 2.03 lies per day ($SD = 3.75$). Lies per day ranged from 0 to 200 (see Figure 2).

Hypothesis H_{1a} predicted that, in aggregate, lie prevalence data is positively skewed with most people lying infrequently and most lies being told by a relatively few prolific liars. For the overall distribution, the *Mode* = 0, *Median* = 1, *Skewness* = 12.74 (skewness for normally distributed data yields a value of approximately 1.00), and *Kurtosis* = 378.16. These values yielded the typical long-tail distribution observed in other lie prevalence studies and were unequivocally consistent with H_{1a}.

Over the three months of the study, no lies were reported for 36.6% of the 57,310 participants-by-day reports ("survey days"); 19.7% were survey days with one lie reported and 16.4% were survey days with two lies reported. Overall, 92.5% of the total survey days involved five or fewer lies. On 4.5% of the survey days, 20 or more lies on a given day were reported.

Alternative lie questions

The central question asked how many times the respondent lied in the preceding 24 h. However, this question was asked in different ways. Most of the participants' reports

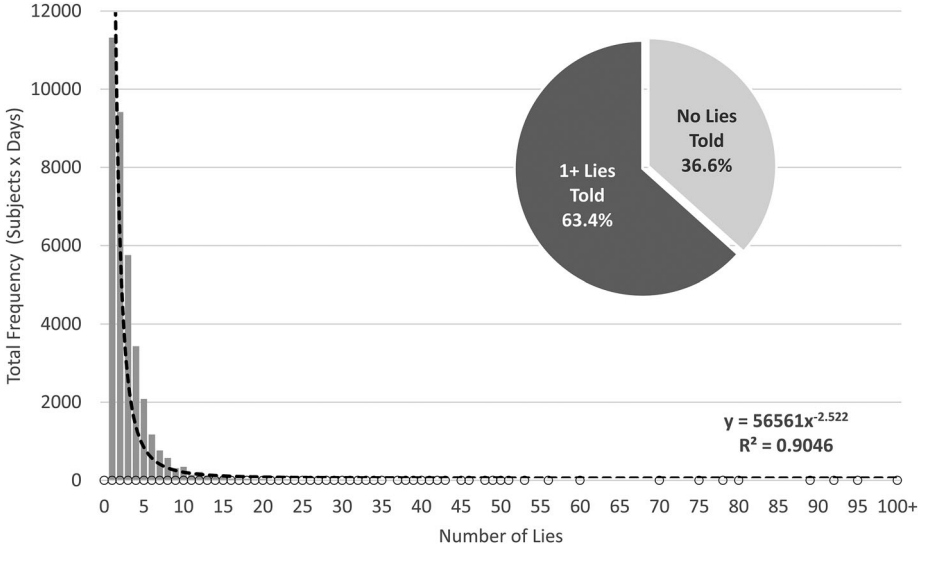

Figure 2. Frequency of daily lies for all participants over three months. Note: The small circles on the x-axis indicate that at least one report of the indicated number of lies.

(n = 50,645; 88.4% of total survey days) were obtained by asking for a single number. The remaining 11.6% of lie reports came from participants who were randomly asked one of four alternative formats that provided additional detail about the nature of the lies told.

Specifically, the first alternative subset was asked about the severity of their lies (using the question from Serota & Levine, 2015); 88.6% of reported lies were described as "little white lies," and 11.4% were characterized as "big lies" (n = 1256). The second subset of respondents was asked a follow-up question regarding what percent of their total communication in the past 24 h consisted of lies (n = 1081). Participants indicated that lies were 7.3% of their total communications, or on average 2.1 lies out of 28.8 total messages per day. Interestingly, this rate is nearly identical to a prior estimate (7.1%, Markowitz & Hancock, 2018). The third subset was asked how the lies were communicated and to whom (n = 3390; based on the question from Serota et al., 2010); 78.9% of the lies were told face-to-face and 21.1% were mediated. Of the lie receivers, 50.6% were friends, 20.8% were family, 11.3% were school/business colleagues, 8.9% were strangers, and 8.5% were casual acquaintances. The fourth subset of reports (n = 938) involved motives to lie. Primary motives reported were: 20.7% to avoid others, 20.0% as humor (a joke or a prank), 14.1% were told to protect one's self, 13.4% were to impress or appear more favorable, 11.0% were to protect another person, 8.7% were for personal benefit or gain, 5.1% were for the benefit of another person, and 1.7% were told in order to hurt another person. The remaining 5.3% of lies were told for unspecified reasons or, explicitly, for no reason at all.

Total lies were computed for each survey day that involved responses from multiple category questions, including white and big lies, who was lied to, and motives for lying. Means and medians were greater when participants are given the opportunity to provide multiple category responses, as intended by the original Serota et al. (2010) question design. A large positive skew, which is the central feature for the comparison of lie distributions, was observed in all five question formats.

Comparison of cross-sectional lie frequency distributions

The current study was designed to look at the distribution of lie behavior for participants over a period of time (within subjects) and observe the distribution of lie behavior across 632 participants on individual days (between subjects). This design resulted in cross-sectional data collection on 91 successive study days or slices.

H_{1b} predicted that, on any given survey day, lie prevalence data is positively skewed with most people lying infrequently and most lies being told by a relatively few prolific liars. Consistent with this prediction, the distribution of total lies for the study (see Figure 2) was replicated on individual days. The mean of the means for the 91 cross-section distributions was the grand mean of the study of 2.03. The average SD for 91 study days was 3.61 (the 91 slices have an average n = 630), but the SD of the daily means (n = 91) was substantially smaller, 0.16 (means $range$ = 1.77–2.66). For 87 of 91 daily slices, the $Median$ = 1 (95.6%) and for 90 of the 91 slices, the $Mode$ = 0 (98.9%). Across the 91 slices, results showed an average $Skewness$ = 9.93 (skewness range = 2.68–19.43) and the average $Kurtosis$ = 92.14 (kurtosis range = 10.63–442.22). The key factor determining variation in distribution shape (represented by the means, SDs, and skews across the 91 cross-sectional studies) was the maximum reported value

(range = 20–200). Constraining the maximum value to 50 lies ($n = 64$ distributions) reduced the average *SD* from 3.61 to 3.06 and the *SD* of the means increased from *SD* = 0.164 ($n = 91$) to *SD* = 0.168 ($n = 64$). In other words, the variation across the 91 cross-sectional slices was negligible, even when the most extreme values were included. Thus, the data were clearly consistent with H_{1b}, and the few prolific liars hypothesis more generally, in each of the daily snapshots. Essentially, this provided 91 non-independent replications.

The daily mean lies varied only slightly from one day to the next (see Online Supplemental Figure 1). On only six of the 91 days was the mean outside the range of the average 95% confidence interval. It is worth noting that this is a student sample and that, over 13 weeks, the peak lie reporting day occurred 12 times on the weekend (Friday to Sunday), suggesting that lie behavior among students is both situational and cyclical. From day-to-day, overall, the average number of lies was relatively stable even with extreme variation in the skewness and kurtosis statistics. Despite variation in the magnitude of the maximum lies reported, H_{1b} is supported by the consistently low median and mode values, further evidence that most participants reported few lies most days, thus accounting for the stability of the trend.

Comparison of individual lie distributions over time

Relevant to H_2, each of the 632 participants reported the number of lies they told each day over the three-month survey period. The distribution of lies for specific individuals over time was the unique feature of this study. Comparing individual-level results required variables that capture their response patterns. For this purpose, descriptive statistics were calculated for each participants' over-time distribution; these statistics – mean, *SD*, median, mode, skewness, kurtosis, minimum and maximum values, and range – were recorded as variables. As with the comparison of cross-section distribution, there was an average distribution: $M = 2.03$ ($SD = 2.879$; *Skewness* = 7.73, $n = 632$). Importantly, although the mean was the same, the *SD* indicated that the variation across participants was substantially larger than the variation of the cross-sectional study days or slices ($SD = 0.164$). Whereas more than 95% of the cross-sectional slices had *Median* = 1, individuals' medians varied widely. Of the 632 participants, 33.2% reported *Median* = 0, 28.0% reported *Median* = 1, and the remaining 38.7% reported a range from *Median* = 2 to *Median* = 35. Both the large *SD* and the wider range of medians point to substantial distribution differences among participants consistent with the idea of a few prolific liars as a stable individual difference.

In addressing RQ_1 about groups of participants reporting similarly shaped distributions of lies over time, a K-Means cluster analysis was performed using the distribution statistics as variables. The data converged quickly. The result was a seven-cluster solution with three major clusters and several "chips," or small numbers of individuals who would not group with the main clusters. Since the non-grouping participants shared some characteristics, two pseudo-clusters were created by collapsing and merging the chips. This reduction yielded five viable clusters.

The vast majority of participants (97.8%) grouped into three of the K-Means clusters. The dominant *normatively low-frequency lie cluster with a tight range* (Cluster 4, "Low-Tight"; $M = 1.33$, $SD = 1.542$, *Skewness* = 1.47, $n_{survey\ \ days} = 43{,}971$; $n_{participants} = 485$,

76.7% of the sample) consisted of participants who were low-frequency liars and who rarely deviated from their low levels of lies/day. The *normatively low-frequency lie cluster with a wide range* (Cluster 3, "Low-Wide"; $M = 1.29$, $SD = 2.282$, *Skewness* = 5.26, $n_{survey\ days} = 3447$; $n_{participants} = 38$, 6.0% of the sample) consisted of participants who were usually infrequent liars but reported a few days with much higher lie/day rates. The *high-frequency lie cluster* (Cluster 2, "High"; $M = 5.17$, $SD = 4.336$, *Skewness* = 1.91, $n_{survey\ days} = 8532$; $n_{participants} = 94$, 14.9% of the sample) consisted of participants whose distributions had a high mean lie frequency and, typically, a wide range of lies/day. The two pseudo-clusters were the *extreme high-frequency lie cluster* (Cluster 1, "Very High"; $M = 17.45$, $SD = 22.817$, *Skewness* = 3.14, $n_{survey\ days} = 542$; $n_{participants} = 6$; 1.0% of the sample) derived from three chips and the *extreme low-frequency lie cluster* (Cluster 5, "Very Low"; $M = 0.12$, $SD = 0.98$, *Skewness* = 20.56, $n_{survey\ days} = 818$; $n_{participants} = 7$; 1.1% of the sample), which included five participants with extremely infrequent lie activity and two with 0 total lies (initially excluded because some statistics could not be computed). Online Supplemental Figure 2 shows the distributions of the five clusters.

The infrequent liar composition of Cluster 5 answered RQ_3 (i.e., Will everyone lie?). Two participants (0.32%) reported not lying at any time over the three-month duration of the study; 99.7% of all participants told at least one lie over a three-month span. Not everyone lied, but the vast majority did; never lying was extremely rare.

In further investigation of RQ_1, to confirm that the clusters are distinct, identify the structure of the data, and plot individual responses, discriminant analysis was performed by using the five clusters as the grouping variable and the descriptive statistics as independent variables. Since 95.7% of the variance was explained by the first two discriminant functions, the analysis was run and then rerun to force a two-dimensional solution. Individual participants and cluster centroids are shown in Online Supplemental Figure 3. Although the discriminant map appears to show participants distributed primarily in the upper-right quadrant, the axes origin is at the geometric center of the distribution and the majority of participants are clustered near the overall sample centroid in the lower-left quadrant. The structure matrix for the two-function discriminant analysis and the pooled within-groups correlations between descriptive statistic variables and the standardized canonical discriminant functions are provided in Online Supplemental Table 1. The independent variables are ordered by the absolute size of correlation within the function. Variables most highly correlated with Function I were kurtosis ($r = .783$) and skewness ($r = .511$). In contrast, Function II was defined by maximum lies ($r = .850$) and range ($r = .830$). Individual participants with high loadings on Function I had over-time lie distributions that tend to be leptokurtic and positively skewed. Individual participants with high loadings on Function II tended to report more days with a high lies/day rate and a wider range of responses. Online Supplemental Table 1 also shows the means and *SDs*, which were calculated for each statistic of the 632 distributions. Online Supplemental Figure 3 shows the two-dimensional plot of all participants and their cluster assignments.

These analyses inform H_{2a} and RQ_2. H_{2a} predicted that participants would vary from day-to-day in the number of lies they tell. Consistent with H_{2a}, whether their lies typically occur with high or low frequency, most participants experienced some bad lie days; that is, a few days with more frequent lying. Further, most frequent liars had days when they

didn't lie often. RQ_2 questioned if the skew observed across individuals would extend to intra-individual distributions. The answer is yes. Although the distributions vary from person-to-person, and clusters vary from other clusters, the values of the skewness statistic were positive across groups and individuals.

Prolific liars versus bad lie days

H_{2a} predicted that people vary from day-to-day in the number of lies they tell, and thus have good and bad lie days. In order to further distinguish between a few prolific liars and those experiencing bad lie days, we examined the combined distributions of the cross-sectional data (between subjects) and the over-time data (within subjects). Figure 3 shows the distribution of 116,366 lies told by 632 participants over 91 days ($M = 2.03$, $SD = 2.88$, $n = 57{,}310$). What this graph elucidates is that most participants told very few lies; only a small portion of the sample told lies with consistently high frequency and most participants reported lies day-to-day within a fairly narrow range of their normative level of lying. In other words, the long-tail distribution was observed both across individuals and over time; some participants lied more than others and most participants had only a few days when they lied more than usual.

In order to parse individual variation across the sample from day-to-day within-person variance, multilevel modeling (HLM; Raudenbush et al., 2011) was used to examine lie days nested within participants. The intraclass correlation was $ICC = 0.583$. Supporting H_{1c}, which predicted individual-level variance in lying, 58.3% of the variance was between subjects and attributable to individual difference that persisted over time. Consistent with H_{2a} the remaining 41.7% of the variance was day-to-day intra-individual variance.

In previous cross-sectional studies, defining prolific liars has focused on the relationship between the shape of the distribution– the long-tail positive skew – and the cumulative lies told (e.g., Serota & Levine, 2015). Lie prevalence data consistently produces a Pareto-like distribution of lies told (excluding reports of zero lies, which cannot be fit to a power function). This indicates that most people tell a few lies and a few people tell most of the lies. The advantage of obtaining lie data with a continuous panel design is the ability to examine lying trends within subjects. If a participant is, in fact, a prolific liar,

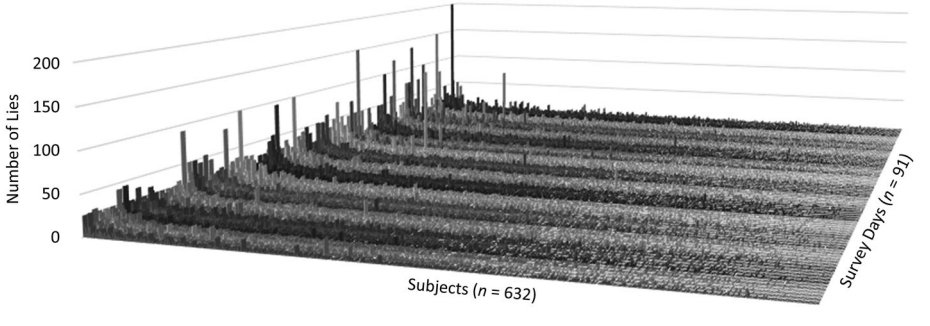

Figure 3. Number of lies by participants by days. Note: 632 participants told 116,366 total lies over 91 days ($M = 2.03$, $SD = 2.88$, $n = 57{,}310$). Participants are ranked by mean lies over 91 days from highest (left) to lowest (right). Survey days are ranked by mean lies/day from lowest (front) to highest (back).

we should be able to observe a pattern of higher frequency lying over time. Once prolific liars are identified and isolated, then among the remaining participants, survey days that positively deviate from normative behavior can be classified as bad lie days. To this end, participants were assigned to one of three categories based on their over-time mean lies. The grand mean of 2.03 lies/day implies that up to two lies is an "acceptable" level for people who are inherently honest. Therefore, participants who averaged 0–2 lies per day were classified as normatively honest. Based on a scree test[2] of the percentage of liars plotted against the percentage of lies told, an average above 5 lies/day appeared to be the natural breakpoint at which lying becomes extensive or prolific for this sample. Therefore, participants who averaged 6+ lies per day were classified as prolific liars. Participants who averaged 3–5 lies/day were classified as intermediate liars, neither distinctly honest nor distinctly prolific. Based on these classifications, normatively honest participants (74.7% of total participants) accounted for 37.8% of total lies told and 65.8% of total survey days consisted of honest participants telling an expected 0–2 lies. Intermediate liars (19.6%) told 34.2% of total lies with 10.0% of survey days consisting of intermediate liars telling 3–5 lies. Prolific liars (5.7%) told 28.0% of total lies and 4.0% of survey days consisted of prolific liars telling 6+ lies (see Table 1). These results provided additional strong support for H_1 *(the few prolific liars hypothesis)*.

Relevant to H_2 and good and bad lie days, the off-diagonal elements of Table 1 represent reported lying above and below category norms. Honest people and intermediate liars had occasional days when they told a lot of lies relative to their baseline (11.7% of survey days); of these, 3.4% were 6+ lies/day or "bad lie days." Prolific and intermediate liars also had days when they told fewer lies than usual (8.5% of survey days); most of these (7.1%) were 0–2 lie days. Confidence intervals ranged from +/− 0.2% points to +/− 0.04% points for largest to smallest subject-day percentages. The presence of proportionally few survey days above the normative range for honest people (more than 2 lies on a given day) and intermediate liars (more than 5 lies on a given day) provided further support for H_2 *(the bad lie day hypothesis)*.

Regarding H_{2b} and the potential misclassification of prolific liars with cross-sectional data, Table 1 suggests that approximately 30% of the time, prolific liars – defined as telling, on average, 6 or more lies per day – reported telling five or fewer lies. Overall, prolific liars accounted for 5.7% of the 57,310 survey days. Of those, 0.3% (5% of prolific liars survey days) were in the 0–2 lies range and 1.4% (25% of the prolific liar survey days) were in the 3–5 lies range. The concern of misclassification was less

Table 1. *Category versus N-Lies on a Survey Day: Percentages of Survey Days and Lies Told.*

	N-lies on a survey day							
	0–2		3–5		6+		Total	
Category Assignment	% Days	% Lies	% Days	% Lies	% Days	% Lies	% Days	% Lies
Honest	65.8%	21.2%	8.3%	14.2%	0.6%	2.4%	74.7%	37.8%
Intermediate	6.8%	4.5%	10.0%	18.6%	2.8%	11.1%	19.6%	34.2%
Prolific	0.3%	0.2%	1.4%	2.8%	4.0%	24.9%	5.7%	28.0%
Total	72.8%	25.9%	19.7%	35.6%	7.5%	38.5%	100.0%	100.0%
n	41,738	30,150	11,271	41,412	4301	44,804	57,310	116,366

Note: Honest participants (M = 0–2 lies) telling 0–2 lies account for 65.8% of all survey days and those lies were 21.2% of all lies told. For intermediate participants, M = 3–5 lies and for prolific participants, M = 6 + lies. Percentages are total table percentages based to total survey days (% Days) and total lies told (% Lies), respectively.

pronounced for normatively honest participants. They were classified correctly on 88% of their survey days and exhibited prolific lying on less than 1% of their survey days.

Self-estimated average of lie frequency

RQ_4 asked if an individual's estimate of their average lie behavior is a good indicator of an individual's normative behavior. Although this study focuses on lie frequency over time, prior to the daily lie survey, participants were asked to estimate how many times they lied during a day *on average*. Participants reported that, on average, they told 3.25 lies per day ($SD = 2.64$; *Skewness* = 1.83). An average lies estimate allows participants to account for variation in their own behavior. These data had a positive skew; however, participants were much less likely to report the extreme values that sometimes occur using the more situationally-specific current-day measure.

The correlation of participants' estimated average lie with their average lie means for the 91 repeated daily lies measures ranged from $r = .17$ to $r = .54$ ($M_r = .39$, $SD_r = .054$; all correlations were significant at $p < .01$, average $n = 606$). A paired-samples t-test for the means of the repeated daily reports ($M = 2.03$) and the average day measure ($M = 3.25$) yielded $t(607) = 10.204$, $CI = 1.141$ to .956, $p < .01$.

Among those reporting 0–2 lies for the single measure average lies/day estimate, 91.8% were classified as honest using the over-time measure (meaning their average daily reports over 91 days were in the 0–2 lies range). Conversely, of those classified as honest using the over-time measures, only 59.2% were correctly predicted by the estimated average lies measure. Of those who estimated telling 6+ lies/day on average, 26.1% were subsequently classified as prolific liars based on their daily reports over three months. But, of those classified as prolific using the 91-day average, 63.9% were predicted by the single average lie/day estimate. Overall, results indicated, for RQ_4, that the single average day measure was only a fair estimate of the mean of repeated daily lie measures, and the extreme values that indicate prolific lying were suppressed.

Discussion

According to Levine (2020), misleading claims asserting the ubiquity of lying and deception are unfortunately and ironically quite common. For example, Pamela Meyer in her TED Talk with more than 31 million views, asserts that "on a given day, studies show that you may have lied to anywhere from 10 to 200 times" (Meyer, 2011). Feldman et al.'s (2002) experimentally induced results are frequently misinterpreted as showing people lie three times in a typical 10-minute conversation. DePaulo et al.'s (1996) classic finding that people lie on average once or twice per day is misunderstood to mean that most people lie at least once every day. It is understandable how people might incorrectly conclude that everyone lies daily. This, however, does not seem to be the case. Our current data even suggest there might be a few people who almost never lie.

In contrast to claims of the omnipresence of deception, the current data suggest that days in which participants reported one or two lies occurred only about one-third of the time (36.1%) and accounted for just 25.9% of the lies. Participants were just as likely to tell no lies on a given day (36.6%). Whereas three or more lies were reported in just one-

quarter of the daily reports, these above average lie days accounted for nearly three-quarters of the 116,366 lies reported.

Summary of current findings

The current study tested predictions from two modules of TDT. Consistent with the few prolific liars module, the positive skew observed in previous research was replicated. Most people lied infrequently and most lies were told by relatively few prolific liars. Extending prior studies, the pattern was shown to hold both within and across survey days. Further, prolific lying was shown to be a pattern that holds for specific individuals over time. Prolific liars are indeed distinguishable from the normatively honest majority, with the caveat that accurately identifying them as such requires examining behavior over time. Indeed, about 60% of the variance in lying is at the individual level.

The study also tested an implication of the deception motive module of TDT which holds that telling specific lies is situationally determined. If the communication situations people experience vary from day-to-day, so too should lying. Inspired by fact-checking data documenting wide daily fluctuations in the false and misleading statements of Donald J. Trump, a good-and-bad-lie-days hypothesis was tested. The data show that people do vary day-to-day. People who are usually honest have days in which they lie more than is typical for them and prolific liars have days in which they tell few lies.

Interestingly, and not previously documented, the day-to-day variance fluctuates considerably from person-to-person. That is, not only are there individual differences in lie frequency, there are identifiable individual differences in distribution shape over time. A sizable majority of our respondents (approximately 75%) did not lie much and were reasonably consistent from day-to-day. For these individuals, zero to two lies per day was typical, and more than four lies per day were two standard deviations out for them. There was however a small group (about 6% of the sample) who had similarly low levels of lying on average, but who had more extreme bad lies days. Generally speaking, prolific liars exhibited much more day-to-day variation than the rest of the sample, this being especially true for the top 1% of liars who had, by far, the largest standard deviations ($SD = 22.8$ compared to 1.5 for the normative majority). The only respondents who did not vary much day-to-day were the 1% who almost never lied. Despite individual differences in both lie frequency and day-to-day variability, the positive skew was ubiquitous. Results for individuals, groups or clusters of individuals, the 91 daily cross-sectional slices, and the aggregate data (combining both individuals and days) all have the characteristic long-tail distribution.

Integration with prior literature

The current findings align well with prior studies of lie prevalence (e.g., Debey et al., 2015; Diaku et al., 2021; Halevy et al., 2013; Markowitz & Hancock, 2018; Park et al., 2021; Serota et al., 2010; Serota & Levine, 2015; Smith et al., 2014), documenting the positive skew in lie behavior and showing that most people are generally honest. The current findings add confidence to prior conclusions showing that the findings of these prior

studies extend beyond cross-sectional data. Along with Markowitz and Levine (2021), the current work also begins to unpack the dispositional and situational precursors of honest and dishonest behavior in tandem.

Critically, the current findings have important implications for the practice of identifying prolific liars with cross-sectional data. In past studies, an individual reporting a high lie count for a single day had been classified as a "prolific liar." Point of fact, these individuals had engaged in "prolific lying" but there was no way of knowing if that single point measurement was indicative of a pattern of behavior. As can be derived from Table 1, nearly half of high-frequency (6+) lie days are reported by non-prolific liars (46%), individuals who averaged five or fewer lies per day over three months. These are bad lie days. Even normatively honest people can be misclassified as prolific. Even so, less than 10% of all 6+ lie days were reported by individuals who tell, on average, 0–2 lies daily. There was also evidence of prolific liars who have good days (on 30% of their survey days they told fewer than 6 lies). Fundamentally, besides providing evidence consistent with our second hypothesis, these findings also reveal that one-time surveys misclassify individuals and are an imprecise method of identifying prolific liars. Cross-sectional data provides an accurate picture of the lie distribution across people but is less well-suited to identifying specific individuals as prolific liars. Asking people to estimate their average lie behavior does not mitigate the problem.

Theoretical implications

This research was situated within TDT and has implications for the theory's empirical adequacy. In contrast to the view that lying is pervasive, TDT holds that (a) lying is infrequent relative to honest communication, (b) most people are honest, (c) the distribution of lying is positively skewed, (d) most lies are told by a few prolific liars, yet (e) the telling of specific lies is situationally determined. The current data align well with all five of these TDT predictions. Consistent with TDT, participants reported that 92.7% of their communication was honest. As with previous data sets, the mode was once again zero lies per day and almost three-quarters (74.7%) of the participants were classified as "honest" with an over-time average of 0–2 lies and their daily values within the 0–2 range nearly two-thirds (65.8%) of the time. Figure 2 shows that the positive skew predicted by TDT is unmistakable. Although those classified as intermediate or prolific liars represented only about one-quarter of the sample, they accounted for nearly two-thirds of the lies. Even more striking, the 5.7% of participants who averaged six or more lies per day told 28% of the total lies. Finally, there was substantial day-to-day variability suggestive of the predicted situational variation. Together, these data add to the empirical support for TDT.

Although several prior studies have examined lie frequency, the current research is the first study to track individual lie behavior daily over a substantial time frame. Doing so was necessary because it allowed us to parse stable individual differences from day-to-day within-person fluctuation. This is theoretically critical because TDT specifies individual differences as well as situationally determined behaviors. Most human interactions are benign but a few are problematic. As people move from situation to situation, lying becomes more or less probable for any given person. The net result is that most relatively

honest individuals will have bad lie days where they lie more than what is usual for them. As a corollary, prolific liars are expected to have good lie days when their lying is not prolific. The current study overcomes the substantial limitation of cross-sectional data, in which individual differences and day-to-day situational variation are confounded.

Findings from the current study add more support to TDT's contention that most people are generally honest. Participants reported that lying is a small portion (less than 8%) of their total communication behavior. Not only do most people tell relatively few lies overall but participants also described almost 90% of their lies as benign "white lies," told without hurtful intent or harmful consequences. Thus, if the overall average is two lies per day and 90% of those lies are inconsequential, then big consequential lies are relatively unusual, perhaps less than 1% of total communication. The very big lie is rare in comparison to most lying in the course of daily life.

Lie prevalence is theoretically important for TDT because it has implications for the utility of the truth-default and truth-bias as well as for deception detection accuracy. According to TDT, people approach most communication from a cognitive mindset where the idea of deception does not come to mind (i.e., the truth-default). Even if suspicion is actively triggered, people are often truth-biased, that is, more prone to evaluate others' communication as honest than otherwise. If lying and deception were pervasive, the truth-default and truth-bias would be dysfunctional, providing a distorted worldview. Alternatively, if deception is as infrequent as TDT predicts, then the truth-default is highly functional and adaptive, a key premise of TDT. Humans are well served by the truth-default, and only need to worry about deception selectively. The so-called truth-bias is not a bias at all but is instead a judgmental tendency that aligns with the base-rate of honesty in everyday communication. The only drawback to truth-default as the normative cognitive state is that it leaves people susceptible to the wily predators among us.

Theoretically, TDT prioritizes lie prevalence as a critical consideration in deception detection. The empirical claim that people are 54% accurate at truth-lie discrimination comes from lab environments where truths and lies are equally represented in the communication environment (Levine, 2020). A major insight of TDT is that to the extent people are truth-biased, recognizing truth-bias improves raw truth-lie discrimination in environments where lies are infrequent. For example, if people passively believed all communication, and if 90% of communication was honest, 9% was little lies, and only 1% constituted consequential lies, then people would have 90% correct truth-lie discrimination and serious duping would only occur about 1% of the time. This paints a very different picture of real-world lie detection than the 54% accuracy finding from experimental work (Levine, 2020) and the misperception that people are being frequently misled.

Beyond TDT, the observation that the positive skew extends to individuals over time may suggest that there is an underlying systemic pattern to lying that is consistent with theories of self-organization in language (Zipf, 1949) and behavioral complexity (e.g., Gleick, 2008). These empirically supported theories show that word usage and semantic content have an underlying order, typically associated with the power function, which reflects the trade-off of efficiency and effectiveness in human behavior. Thus, as our data showed, the Trump pattern was widely repeated across individuals, albeit with expected lower frequencies. Skewed distributions observed previously in cross-sectional

studies may be as likely to occur because of the aggregation of individuals' varying experiences on a given day as they are to reflect individual differences in lie behavior. It is likely that both individual variation and individual differences have been captured by previous research, but due to obtaining only single-point-in-time measurements we have not previously been able to separate the two.

College students and adults

A common source of fascination is in the similarities and differences between college student samples and representative adult samples. As mentioned in the introduction, the general trend is that lying increases with age through childhood, peaks during the teen years, and then declines gradually with age. Despite maturation, the positive skew is maintained (Debey et al., 2015; Serota et al., 2010). Understanding the size and nature of these trends, however, requires considerable nuance. The difference was the largest in the original DePaulo et al. (1996) study where adults lied once per day and students twice per day. Depending on how one frames the results, this is a small difference of one lie per day or a difference of 100% with students lying twice as often as adults. If we rely on the much larger samples in the current study compared to Serota et al. (2010), the mean difference is more modest; 2.03 for students compared to 1.65 for adults.

However, if there is a take-away message from this article, it is that averages are misleading when data are highly skewed, and the prevalence of lying is most definitely skewed. Means are impacted by "outliars" (i.e., the extreme lying scores). The difference between college students and adults appears to be at the low end of distribution. In Serota et al. (2010), almost 60% of adults reported telling no lies on the day of the survey whereas the current investigation found a rate almost half that (36.6%). DePaulo et al. (1996) report a similar, albeit stronger, pattern. Thus, the key difference between students and adults seems to be in the proportions of participants telling zero, one, and two lies per day. The distribution within the normatively honest group shifts. The skew and the few prolific liars findings generalize across samples.

Limitations

Readers new to the lie prevalence literature are often concerned about the validity of self-report measures such as those used in the current research. Lying is socially disapproved, so social desirability is a seemingly obvious concern. Readers will wonder how we know that the participants aren't lying about how often they lie. Surely, they under-report either for self-presentation or self-concept motivations. However, if TDT and our current findings are correct, then most people are honest. This applies to filling out research surveys, too. To check this, Halevy et al. (2013) performed a behavioral validation of self-report lie frequency measures. Self-reported honesty, in the form of infrequent self-reported lying, predicted observed honesty in experimental integrity challenges.

The biggest limitation of the current research is also ironically its biggest strength. This was a massive and arduous undertaking. TDT is built around replication, but the study was logistically challenging in ways that might discourage efforts to replicate this work. Although the design and sheer amount of data are assets, the practical obstacles

to replication are not. Nevertheless, we hope that our current findings demonstrate the value of longitudinal data and that other researchers follow our lead and put forth the effort.

There are a variety of other limitations worth mentioning. Although we provided a definition of lying, that definition might not have been evenly applied across participants. Consequently, some of the individual-level variation may reflect a different definition rather than a difference in actual behavior. Further, individuals' recall was surely far from perfect and participants may have been less likely to remember little white lies.

Some may find our terminology of good and bad lie days confusing, although this is a limitation of the report and not the research. We use bad lie days to refer to quantity; days on which people tell an unusually high number of lies. Of course, bad lie days could also refer to days on which unusually big bad lies are told. We think, however, our idea holds for both uses of the phrase. Just as frequent lie days are unusual relative to infrequent lie days, our data show that big bad lies are infrequent relative to little white lies.

Another issue of concern is external validity and the generality of the results. One might ask, for example, "How do you know that six or more lies is prolific lying?" The answer is we don't. What we do know is that for almost all self-reported lie studies, regardless of sample composition – age, sex, nationality – the majority of respondents are normatively honest, typically with a median of one or two lies and a mode of zero. But as this study shows, the tendency to tell more lies than normal is a combination of situational influences and individual differences. Based on previous research, we know that some groups of individuals generally tell more lies; for example, college students and especially teenagers (Debey et al., 2015; Levine et al., 2013; Serota & Levine, 2015). However, the average is not the primary metric for classifying lie behavior; rather it is where the individual response falls within a distribution. These studies have all exhibited results with a short head, a long tail, and an ambiguous middle. The tail may become distinct at four, five, six, or more lies in a 24-hour period. The generalization is that, regardless of sample, each of these studies reveals a small number of people (or days, for individuals observed over time) that account for a disproportionally large share of the lies told.

Future directions

The current findings have implications for research seeking to link lie behavior with specific personality traits or demographic characteristics (e.g., Michels et al., 2020). As we noted in the introduction, such approaches produce inconsistent findings and have had limited success predicting who will lie. If the current findings are accurate and projectable, then perhaps only 60% of the variance in lie behavior is potentially explainable in terms of individual difference predictor variables. However, we only know this because individuals were tracked daily over three months. On any given day, a person's behavior may reflect either their dispositions or their situational good or bad lie days (or both). Thus, studies using cross-sectional data to classify prolific liars are shown here to risk the misclassification of individuals.

More generally, future research needs to further unpack the interplay of individual differences, situational features, and specific deception motives. Presumably, individual

differences such as demographics, occupation, and personality lead people to experience different situations where the truth will be more or less consistent with communication goals. Personality is further likely to impact the priority individuals place on achieving different communication goals, and a person's willingness to use deceptive means for goal attainment. Linking how often people lie (prevalence) with why people lie (motives) is fertile ground for future research.

Conclusion

The present study extends previous research on lie prevalence by collecting reports of daily lies over three months with a continuous panel that allows the parsing of individual patterns from day-to-day variability. Two hypotheses from TDT were tested. Compelling evidence for the presence of both a few prolific liars and bad lie days were obtained. Previous findings that most communication is honest, most lies are inconsequential little white lies, the distribution of lying is skewed, and that most lies are told by a relatively few prolific liars were all successfully shown to hold over time. Notably, considerable day-to-day variation was observed, as were large individual differences in day-to-day variation. The findings confirm that prolific liars are distinct and identifiable people but also demonstrate prolific lying is not isomorphic with being a prolific liar. On any given day not all high-frequency liars are prolific, and those who are prolific do not always exhibit prolific lying. Observations of extensive lying on a single day only indicate a prolific liar about one time out of four. These findings add support for TDT and reinforce that lie detection is a low base-rate activity. Above all, findings from the current study document that for most people lying is less prevalent than often believed.

Notes

1. We recognize that false and misleading statements are not isomorphic with lies. False statements are not lies if the sender believes them. Similarly, not all lies can be fact checked. Nevertheless, considering false and misleading statements as a proxy for lies leads to insights of potential theoretical importance that motivated the current study.
2. Serota et al. (2010) and Serota and Levine (2015) created rules for identifying the breakpoint for prolific lying based on fitting data to Poisson and Pareto distributions. Although the scree test was originally developed for identifying significant eigenvalues in a factor structure, it can be used to delineate the "elbow" in any power law distribution. While subjective, the scree test is more efficient than applying the Poisson/Pareto rules and typically yields a similar result.

Acknowledgements

The authors wish to thank Hee Sun Park (Korea University) for assistance with the HLM analysis, the technical support staff of Qualtrics for design and implementation guidance, the faculty of the Department of Communication Studies, University of Wisconsin-La Crosse for participating in the recruitment of subjects and for providing course credit as an incentive, Christopher J. Carpenter (Western Illinois University) for review of data, and the three anonymous reviewers for their helpful feedback and guidance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

ORCID

Kim B. Serota http://orcid.org/0000-0001-7870-2633
Tony Docan-Morgan http://orcid.org/0000-0002-2925-654X

References

Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior Pinocchio: A cross-sectional lifespan investigation of deception. *Acta Psychologica*, *160*, 58–68. https://doi.org/10.1016/j.actpsy.2015.06.007

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. https://doi.org/10.1037/0022-3514.70.5.979

Diaku, 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

Docan-Morgan, T. (2019). *The Palgrave handbook of deceptive communication*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-96334-1.

Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology*, *24*(2), 163–170. https://doi.org/10.1207/S15324834BASP2402_8

George, J. F., & Robb, A. (2008). Deception and computer-mediated communication in daily life. *Communication Reports*, *21*(2), 92–103. https://doi.org/10.1080/08934210802298108

Gleick, J. (2008). *Chaos: Making a new science*. Penguin Books.

Halevy, R., Shalvi, S., & Verschuere, B. (2013). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research*, *40*(1), 54–72. https://doi.org/10.1111/hcre.12019

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. (2020). *Duped: Truth-default theory and the social science of lying and deception*. University of Alabama Press. https://doi.org/10.1093/joc/jqaa026.

Levine, T. R., Ali, M. V., Dean, M., Abdulla, R. A., & Garcia-Ruano, K. (2016). Toward a pan-cultural typology of deception motives. *Journal of Intercultural Communication Research*, *45*(1), 1–12. https://doi.org/10.1080/17475759.2015.1137079

Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). 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., Punyanunt-Carter, N. M., & Moore, A. (2020). The truth-default and video clips: Testing the limits of credulity. *Communication Studies*, *72*(2), 133–145. https://doi.org/10.1080/10510974.2020.1833357

Levine, T. R., Serota, K. B., Carey, F., & Messer, D. (2013). Teenagers lie a lot: A further investigation into the prevalence of lying. *Communication Research Reports*, *30*(3), 211–220. https://doi.org/10.1080/08824096.2013.806254

Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*(6), 895–919. https://doi.org/10.1037/0033-2909.132.6.895

Markowitz, D. M., & Hancock, J. T. (2018). Deception in mobile dating conversations. *Journal of Communication*, *68*(3), 547–569. https://doi.org/10.1093/joc/jqy019

Markowitz, D. M., & Levine, T. R. (2021). It's the situation and your disposition: A test of two honesty hypotheses. *Social Psychological and Personality Science*, *12*(2), 213–224. https://doi.org/10.1177/1948550619898976

Meyer, P. (2011, July). How to spot a liar [Video]. TED Conferences. https://www.ted.com/talks/pamela_meyer_how_to_spot_a_liar.

Michels, M., Molz, G., & Maas genannt Bermpohl, F. (2020). The ability to lie and its relations to the dark triad and general intelligence. *Personality and Individual Differences*, *166*, 110195. https://doi.org/10.1016/j.paid.2020.110195

Palomares, N. A. (2014). The goal construct in interpersonal communication. In C. R. Berger (Ed.), *Interpersonal communication* (pp. 77–99). De Gruyter Mouton. https://doi.org/10.1515/9783110276794.77.

Park, H. S., Serota, K. B., & Levine, T. R. (2021). In search of Korean outliars: "A few prolific liars" in South Korea. *Communication Research Reports*, *38*(3), 206–215. https://doi.org/10.1080/08824096.2021.1922374

Raudenbush, S. W., Bryk, A. S., Cheong, A. S., Fai, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Scientific Software International.

Serota, K. B. (2019, April 3–7). Our pathological, prolific liar in chief: The crisis of Donald Trump's lies. In F. Yang (Chair), *The crossroads at "truth isn't truth": using communication theory to make truth great again*. Symposium conducted at the Southern States Communication Association 89th Annual Conference, Montgomery, AL, United States. https://www.ssca.net/2019-convention.

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), 138–157. https://doi.org/10.1177/0261927X14528804

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

Smith, M. E., Hancock, J. T., Reynolds, L., & Birnholtz, J. (2014). Everyday deception or a few prolific liars? The prevalence of lies in text messaging. *Computers in Human Behavior*, *41*, 220–227. https://doi.org/10.1016/j.chb.2014.05.032

Stockman, D. (2017). *Psychopathie en deceptie: De relatie tussen psychopathie, sociale wenselijkheid en leugenachtig gedrag.* [Unpublished master's thesis]. University of Ghent. https://doi.org/10.13140/RG.2.2.30017.25440.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Zvi, L., & Elaad, E. (2018). Correlates of narcissism, self-reported lies, and self-assessed abilities to tell and detect lies, tell truths, and believe others. *Journal of Investigative Psychology and Offender Profiling*, *15*(3), 271–286. https://doi.org/10.1002/jip.1511