Individuals with depression are prone to maladaptive patterns of thinking, known as cognitive distortions, whereby they think about themselves, the world, and the future in overly negative and inaccurate ways. These distortions are associated with marked changes in an individual’s mood, behavior, and language. We hypothesize that societies can undergo similar changes in their collective psychology that are reflected in historical records of language use. Here, we investigate the prevalence of textual markers of cognitive distortions in over 14 million books for the past 125 y and observe a surge of their prevalence since the 1980s, to levels exceeding those of the Great Depression and both World Wars. This pattern does not seem to be driven by changes in word meaning, publishing and writing standards, or the Google Books sample. Our results suggest a recent societal shift toward language associated with cognitive distortions and internalizing disorders.

Depression is a leading contributor to the burden of disability worldwide (1, 2). Some evidence indicates that disability attributed to depression has been rising over the past decades, particularly among youth (3–5). Can societies collectively become more or less depressed over time, as their populations are exposed to stressors such as war, political upheaval, economic collapse, food insecurity, inequality, and disease (6, 7)? This question is difficult to answer for long time scales because formal diagnostic criteria were introduced only 40 y ago and these criteria have undergone changes over time (8).

Depression is associated with distinct and recognizable maladaptive thinking patterns, referred to as cognitive distortions, wherein individuals think about themselves, the future, and the world in inaccurate and overly negative ways (9–12). For example, a cognitive distortion seen in depression occurs when individuals label themselves in negative, absolutist terms (e.g., “I am a loser”). They may talk about future events in dichotomous, extreme terms (e.g., “My meeting will be a complete disaster”) or make unfounded assumptions about someone else’s state of mind (e.g., “Everybody will think that I am a failure”). Typologies of cognitive distortions generally differentiate between a number of partially overlapping types, such as “catastrophizing,” “dichotomous reasoning,” “disqualifying the positive,” “emotional reasoning,” “fortune telling,” “labeling and mislabeling,” “magnification and minimization,” “mental filtering,” “mindreading,” “overgeneralizing,” “personalizing,” and “should statements.”

The theory underlying cognitive-behavioral therapy (CBT), the gold standard for the treatment of depression and other internalizing disorders (13), holds that cognitive distortions are associated with internalizing disorders; they reflect negative affectivity and avoidant behavioral patterns in the context of environmental stress (14, 15). Language is closely intertwined with this dynamic. In fact, recent research shows that individual
disorders such as depression and anxiety, in millions of books published over the course of the last two centuries in English, Spanish, and German. We find a pronounced “hockey stick” pattern: Over the past two decades the textual analogs of cognitive distortions surged well above historical levels, including those of World War I and II, after declining or stabilizing for most of the 20th century. Our results point to the possibility that recent socioeconomic changes, new technology, and social media are associated with a surge of cognitive distortions.
Fig. 1. Examples of CDS n-grams shown inside gray boxes, surrounded by plausible context words that may vary without affecting whether the n-gram marks the expression of a cognitive distortion of the given type (e.g., mindreading, emotional reasoning, or labeling and mislabeling). CDS were designed by a team of CBT experts, linguists, and native language speakers to capture the expression of a particular cognitive distortion type, regardless of its specific lexical context. For English (US), Spanish, and German the team of experts defined respectively 241, 435, and 296 n-grams to mark 12 commonly distinguished types of cognitive distortions. Note that our prevalence measurements count only the CDS n-gram occurrence regardless of context (“everyone thinks,” “still feels,” and “I am a”). A complete list of all CDS n-grams by distortion type is provided in SI Appendix, Tables S1–S3.

Results

We perform this analysis for three unique geographic and linguistic spheres: 1) the United States of America (US English), 2) the German-speaking countries (German), and 3) all Spanish-speaking countries (Spanish). English (US), Spanish, and German were chosen as the focal points of our analysis because they share a common alphabet, a common history and vary in terms of whether they are spoken as a first language either in a particular geographic region (US books only and German-speaking countries) or across several continents (Spanish) as a control. We limit our analysis to the range 1855 to 2019 since it provides 125 y of persistently high publishing volume for all three languages and few grammatical, orthographic, or spelling changes that would affect our analysis. Although books published in a specific language in a particular geographic area are not necessarily a representative reflection of society as a whole, persistent language trends over decades and centuries, observed from tens of millions of books, have in previous research been shown to signal cultural, linguistic, and psychological changes (18, 20–28).

Trends for English (US), Spanish, and German. We first examine the history of the median prevalence (z scores) of the entire set of English cognitive distortion schemata (n = 241) in English (US) books (N = 9,018,119, United States only), from 1855 to 2019 (Fig. 2A). Since these data pertain only to books published in the United States, we mark notable events in US history or notable changes in the time series: the end of the century in 1899; the start of World War I; the financial collapse of 1929; the start of World War II; a peak of CDS prevalence in 1968; and distinct trend changes in 1978, 1999, and 2007.

The overall trend of CDS prevalence for most of the 20th century pointed distinctly downward toward a historic minimum in 1978, with only a few noticeable peaks, one surrounding the turn of the century in 1899 (possibly related to the Spanish–American war), a slight peak from 1940 to 1945 (around the time of World War II), and a sharp peak in 1968 (possibly related to social and political unrest). From 1978, we observe an accelerating increase in CDS prevalence. This acceleration seems to be separated into three periods: an accelerating increase from 1978 to 1999 (where CDS prevalence first exceeds levels observed in the 1910s), an even more rapid increase after 1999 to roughly 2007, followed by an acceleration after 2007, and a possible stabilization in 2010. The so-called “bursting of the dot.com bubble” seems to coincide with an acceleration of the increase of CDS prevalence after 1999 whereas the acceleration since 2007 seems to coincide with the widespread uptake of social media and the start of the Great Recession. Present CDS prevalence levels exceed those observed since the 1900s by almost two standard deviations (excepting the 1899 peak).

We include Spanish in our analysis as a control vs. English (US) and German since it is not confined to a particular geographic region; i.e., these data encompass all books published in Spanish, which includes Spain (Europe) and most of Latin America (N = 1,658,438 books). The prevalence of Spanish CDS markers (N = 435 n-grams) remains quite stable throughout the 20th century, with a moderate increase around the start of World War I, a very short moderate spike in 1929, and an upward
departure from a 30-yr downward trend in 1953 after which levels seem to level off until 1984 (Fig. 2F). Starting in 1984, however, we observe the same hockey-stick pattern as we saw for English (US): a sharp acceleration of an upward trend starting in 1984 leading to present CDS prevalence levels that exceed the historical baseline by more than one standard deviation. This trend seems to accelerate in 2008.

The pattern of prevalence for German (Fig. 2C) books (N = 3,843,962) provides face validity for the ability of our CDS markers (N = 296 n-grams) to capture significant moments of stress in a population, since they match major historical and geopolitical events that are specific to German-speaking countries (predominantly Germany and Austria). Contrary to English and Spanish, CDS prevalence levels start relatively low around the 1900s, but sharply increase since the start of World War I, reaching a peak in 1920 and 1923, coinciding with the aftermath of World War I in Germany and a devastating recession in 1923. Throughout the existence of the Weimar Republic, we observe decreasing levels of CDS markers.

However, this trend is interrupted in 1932 at which point cognitive distortion levels increase sharply. This period includes major social upheaval, economic struggles, the end of the Weimar Republic, the emergence of the Nazi regime, and the start of World War II. CDS prevalence levels increase rapidly during World War II, reaching their peak in 1946, the year after Germany was defeated. CDS prevalence declines sharply afterward and reaches a stable plateau throughout the 1950s to 2007, with only a minor peak in 1962 and no indications of accelerating CDS prevalence levels during the 1970s or 1980s as we observe for English (US) and Spanish. Notably, in 2007, at the start of the worldwide Great Recession we see a nearly immediate increase of CDS prevalence levels to nearly two standard deviations above the historical mean.

**Null Model of Randomly Chosen N-Grams.** We compare the pattern of change for all three languages in terms of the 95% confidence intervals of CDS prevalence (Materials and Methods, Bootstrapping) for English (US), Spanish, and German against a null model of 10,000 sets of 241 randomly chosen n-grams (Fig. 3). These sets of random n-grams were sampled from all n-grams in the respective CDS set and the same bias toward recently published books due to increased publication volume over time (Materials and Methods, Null Model).

We provide annotations that mark significant historical events in the graph that have affected the three populations such as the financial crisis of 1929 (“Wall St. crash”), the two World Wars, and the great recession starting in 2007. CDS prevalence levels for English (US), Spanish, and German significantly exceed those of this null model in recent decades, but in the case of Germany also during World War I and World War II. Note that English (US) levels fall below those of the null model from the 1920s to the 1990s (SI Appendix, Fig. S6).

**Trends for Distinct Cognitive Distortion Types.** We plot the time series of yearly mean CDS prevalence separated by 12 commonly recognized cognitive distortion types (14) for English (US), Spanish, and German (Fig. 4). For all three languages and across most cognitive distortion types we see the characteristic hockey-stick signature of stable or declining CDS prevalence levels followed by a surge above historical levels during the period 1980 to 2010 to levels well above the historical mean. The one exception is statesmen, which, due to their grammatical structure, may be difficult to translate to n-grams uniquely associated with the specific cognitive distortion.

For English we frequently see a “tilted hockey stick” pattern where certain types of CDS n-grams declined over the 20th century followed by a rapid surge of prevalence since 1978. This is the case for fortune telling, overgeneralizing, magnification and minimization, mindreading, and labeling and mislabeling and perhaps most pronounced for dichotomous reasoning, suggesting that these distortion types are likely responsible for the decline of CDS prevalence observed throughout the 20th century (Fig. 2A). We also find slight peaks for catastrophizing, emotional reasoning, and mindreading at the time of the US involvement in World War II. For German, however, we see peaks surrounding World War I and World War II for dichotomous reasoning, fortune telling, labeling and mislabeling, mental filtering, mindreading, overgeneralizing, personalizing, and should statements, possibly indicating the widespread effects of the two World Wars on German language use.

**Robustness and Limitations.** As our observations could be caused by a number of effects and biases, we conducted a number of mitigation controls and sensitivity analyses to test alternative explanations for the observed patterns.

**Language effects.** We caution that changes in meaning or semantic shift of the CDS n-grams may potentially bias our results. As a rhetorical example, the 1-gram “ever” could acquire a different meaning or usage over time and hence lose its meaning as a marker of a cognitive distortion of the dichotomizing
type. We perform several controls to account for changes in language over time. First, the CDS n-grams consist predominantly of words that have been among the most frequent since 1895 [mean word percentile among all 1-grams \( M(P_w) = 99.885, SD = 0.346; SI Appendix, Fig. S1 \)]. The CDS n-grams have equally been among the most frequent since 1895 [mean CDS n-gram percentile among all 2- to 5-grams \( M(P_c) = 0.946, SD = 0.010; SI Appendix, Fig. S2 \)]. Hamilton, Leskovec, and Jurafsky (25) quantify semantic shifts over historical time using word embeddings, showing that frequent words experience the lowest rate of change, scaling with an inverse power law of word frequency. Hence the rate of semantic shift of the words in our CDS n-grams and the CDS n-grams themselves could be among the lowest as well. Second, a trend toward shorter sentences (29) may provide alternative explanations of our observations, but although sentence length did decrease from 1890 to the 1920s in English, it has remained stable since (30). Furthermore, our analysis accounts for changes in sentence length by normalizing n-gram prevalence by the frequency of end-of-sentence punctuation for that year (Materials and Methods, Time Series: Prevalence and Normalization). Finally, we previously showed that the prevalence of the CDS n-grams in the language of individuals with depression is not affected by the emotional valence of the n-grams or the presence of personal pronouns (17); hence, a language trend toward more emotional language of individuals with depression is not affected by the emotional valence of the n-grams. However, we observe the opposite, a significant increase of CDS prevalence.

Second, the choice of CDS n-grams could lead to a “recency bias” in our results, explaining their rise in prevalence in recent decades. We control for this effect with a null model that samples random n-grams more frequently from recent books, due to rapidly increasing publication volume since 1895, thereby inducing a bias toward more recent language. We observe increases of CDS n-gram prevalence well above levels predicted by this null model (Fig. 3). Hence, a recency bias alone may not likely explain the observed surge in CDS prevalence in recent decades relative to this null model.

Finally, all n-grams in the English (US), Spanish, and German CDS sets occurred in every year from 1895 to 2019, indicating they were in continuous use throughout this period. They were highly frequent from 1895 to 2019, in fact on average more frequent than 94.6% (SD = 0.0103) of all n-grams in the Google Books data (SI Appendix, Figs. S1 and S2). We furthermore bootstrap our prevalence estimates to gauge the sensitivity of our findings to random changes in the set of CDS n-grams over time (Materials and Methods, Bootstrapping). The narrow 95% CI bands (Fig. 3) throughout the period under consideration indicate the stability of our observations over time.

**CDS limitations.** We caution that although the Google Books data have been widely used to assess cultural and linguistic shifts, and they are one of the largest records of historical literature, it remains uncertain whether CDS prevalence truly reflects changes in societal language and societal wellbeing. Many books included in the Google Books sample were published at times or locations marked by reduced freedom of expression, widespread propaganda, social stigma, and cultural as well as socioeconomic inequities that may reduce access to the literature, potentially reducing its ability to reflect societal changes. Although CDS n-gram prevalence was shown to be higher in individuals with depression (17) and our composition of CDS n-grams closely follows the framework of cognitive distortions established by Beck (9), they do not constitute an individual diagnostic criterion with respect to authors, readers, and the general public. It is also not clear whether the mental health status of authors provides a true reflection of societal changes or whether cultural changes may have taken place that altered the association between mental health, cognitive distortions, and their expression in language.

**Discussion**

While the differences between the languages are interesting, perhaps the most important point is that the expression of cognitive distortions increases for all three languages in the recent three decades, leading to a distinct hockey-stick pattern indicating a surge of the CDS prevalence levels, which serve as lexical markers of cognitive distortions.

We can only speculate on the possible underlying causes of the observed surge of CDS prevalence for these three languages, since our results do not establish any causal mechanisms. The
strong increases of CDS prevalence in German during World Wars I and II are validating with respect to the ability of our CDS n-grams to signal societal dynamics in times of turmoil and run counter to the hypothesis that our results are caused by a recency bias in our choice of CDS n-grams and the Google Books sample. In fact, the surge of CDS prevalence during and right after World War II may be the product of a detrimental combination of the war experience and National Socialist propaganda. While there was not a separate language of National Socialism (32), the discourse of National Socialism invaded many registers of speech, including everyday language use, and thus normalized the political agenda (33, 34). In particular, the discourse of National Socialism is shaped by a language of identity that emphasizes an us-them divide (35), which relates to several markers of CDS n-grams, e.g., dichotomous reasoning, labeling and mislabeling, mindreading, and fortune telling. The tumultuous period of 1959 to 1962 in Germany cemented the division of East and West when the Berlin Wall was built in 1961. Interestingly, the fall of the Wall in 1989 did not result in higher amounts of psychological distress (36) and also does not register on fluctuations of CDS n-grams prevalence. The German data show a period of stability from 1962 up to the rapid increase after 2007.

Other differences between the dynamics of cognitive distortions in the three language corpora we analyzed might also point to relevant drivers. For instance, in Spanish and English we see a rising trend starting around 1980, whereas in German there is no such rise, only a sharp jump to a higher level in 2007. It is possible that the reunification of Germany in 1990 and the increased integration of the German-speaking countries in the European Union (and the introduction of the Euro currency in 1999) provided resilience to trends recorded in Spanish and English prior to 2007.

It is suggestive that the timing of the US surge in CDS prevalence coincides with the late 1970s when wages stopped tracking increasing work productivity. This trend was associated with rising income inequality to recent levels not seen since the 1930s (37). This phenomenon has been observed for most developed economies, including Germany, Spain, and Latin America, contemporaneous with the rapid growth of automation and demand for highly skilled labor (38). The great recession of 2007 might have compounded the effects of this decades-long trend that started in the late 1970s. The widespread adoption of communication technologies such as the internet, the World Wide Web, and social media (39–42) may have driven greater societal and political polarization (43, 44) at a global level (45). The language of such polarization may correspond to cognitive distortions (46), in particular us-vs.-them thinking (labeling and mislabeling), dichotomous reasoning, mindreading (47), overgeneralizing, emotional reasoning, and catastrophizing.

We caution that we make no causal claims with respect to the relationship between lexical markers, cognitive distortions, and internalizing disorders, and the above comments therefore constitute speculations that we hope may inspire follow-up research. Regardless of speculation with respect to the underlying cultural, social, or economic drivers, our results indicate historically high levels of the expression of a large set of lexical markers of cognitive distortions in three languages. Given the association between cognitive distortions and internalizing disorders, this points to the possibility that large populations are increasingly stressed by pervasive cultural, economic, and social changes. The rising prevalence of depression and anxiety (3–5) in recent decades seems to align with our observations.

The availability of large-scale historical records of published languages going back centuries may provide a unique opportunity for the quantitative investigation of important cultural and linguistic dynamics (“culturometrics”) (21), while acknowledging limitations with respect to verifying hypotheses and testing the causal mechanisms that underlie any observations from these data. Future work may contribute to a better understanding of how changes in the collective psychology of societies can be observed over time and how these changes are manifested in their language in response to a variety of cultural and socio-economic challenges, for example from quantitative indicators of semantic shifts in word meaning (25).

Materials and Methods

Google Books N-Gram Data. We used the third version 2019 release of the Google Books n-gram data, which the Google Books team makes freely available online (https://storage.googleapis.com/books/ngrams/books/datasetsv3.html). The data span from the 16th century to the year 2019, with increasing coverage of later years as publication volume grew rapidly.

Cognitive Distortion N-Grams. A panel of CBT experts engaged in a collaborative design effort to compile a set of 241 English n-grams (sequences of n = 1, 2, 3, 4, and 5 words) that were deemed to indicate the expression of a cognitive distortion of a particular type. These CDS n-grams were designed to consist of simple, frequent, and nontechnical expressions, e.g., “I am a,” “he thinks,” “will be,” etc., designed to be stand-alone expressions of cognitive distortions regardless of their specific context. For example, “I am a [defeated gentlemen]” and “I am a [loser]” both constitute an expression of a labeling and mislabeling captured by the “I am a” 3-gram.

The set of 241 English n-grams was subsequently translated from English into Spanish and German. This translation mainly focused on retaining the cognitive distortion expression of an n-gram in the target language. All translations were collaboratively compared, back translated, and validated by members of our team of CBT experts and native language speakers to ensure consensus. Since CDS n-grams are short expressions of frequent one, two, three, four, or five words, many have nearly literal translations, e.g., “I am a” translates to “Yo soy un” and “yo soy una.” We provide the complete lists of all English, Spanish, and German CDS n-grams in SI Appendix, Tables S1–S3.

The number of CDS n-grams is higher in Spanish and German than in English since the former need to capture grammatical variations such as conjugations, cases and inflections, and gender. Some n-grams were translated to regular expressions (RE) to capture succinctly all possible lexical and grammatical variations of the same CDS in Spanish and German. For example, the English “I am a” was translated to the REs “Yo soy un(a),” which matches both the male and female gender of the definite article in Spanish, and “ich bin ein(e),es,em,er,es” to match all possible grammatical variations in German (including misspellings or errors). Note that in the case of German the operative verb or term is frequently at the end of the sentence. For example, for the n-gram “I never” was translated to the regular expression “ich (+) nie,” matching any 3-, 4-, and 5-gram that starts with “Ich” (I) and ends in “nie” (never) as “Ich habe nie” (I have never), “Ich hatte nie” (I never had), and “Ich war nie” (I never was). All n-grams and REs were matched in a case-insensitive manner against the Google Books data to capture the widest possible lexical variation across time, including all capitalizations. One given CDS, depending on capitalization and grammatical variations, could match multiple expressions. For example, the German regular expression “Ich bin ein(e),es,em,...” could match any of “ich bin ein,” “Ich bin eine,” “Ich bin eines”, etc., including some variations that are rare or even grammatically incorrect (e.g., other letters capitalized). All such matches for each individual n-gram were summed into a singular compound time series for the specific CDS such that the most frequent forms have the highest relative weight in subsequent analysis.

Each match retrieved the complete time series of n-gram frequencies for the specific n-gram and matching RE from the earliest to the latest year provided. The time series values correspond to the number of times the n-grams occurred in the Google Books data in a particular year. Note that the earliest books in the Google Books dataset were published in the 15th century, a period marked by low publication volumes and variances in spelling and grammar. As mentioned, our analysis was limited to the period 1895 to 2019 to capture the end of the 19th century, most of the 20th century, and the past two decades, a period of high publishing volume, relatively stable orthographic standards, and relatively low variance of our CDS prevalence data (SI Appendix, Fig. S4). Each n-gram in the CDS set of each language occurred in every year from 1895 to 2019, indicating continuous coverage for all individual CDS n-grams.
Time Series: Prevalence and Normalization. The volume of books published has increased significantly over the past two centuries, punctuated by declines at times of economic collapse and war (SI Appendix, Fig. S3). The frequency of occurrence of any specific n-gram will therefore fluctuate accordingly, since it is recorded from a changing sample of books published. We therefore determine yearly n-gram prevalence by normalizing the observed yearly frequency of each n-gram by the total yearly volume published. We estimate the latter by summing the yearly frequency of periods, exclamation points, and question marks (“,” “!”,” and “?”), three punctuation symbols that are used in English, Spanish, and German to mark the end of a sentence. Their frequency indicates publication volume as the number of sentences published (SI Appendix, Fig. S4), accounting as the number of sentences published (\( M = 0.9633, SD = 0.0103 \)), with periods approximately 26 times more frequent than exclamation points and question marks.

More formally we determine our prevalence time series as follows. We define a set \( C_\text{n} \) of \( k \) n-grams \( c \in C_\text{n} = \{ c_1, c_2, \ldots , c_j \} \), where the number of words in each n-gram is its length \( n \in (1, 2, 3, 4, 5) \). We denote the yearly time series of the prevalence of any n-gram \( c \) as the set \( X(c_j) \), where \( j \) refers to a year.

\[
X(c_j) = \frac{\sum_i c_i}{\sum_i N_i} \quad \forall \ c_j \in C_\text{n}.
\]

Each value \( X(c_j) \) of \( X(c_j) \) represents the n-gram's prevalence, i.e., the total number of occurrences of n-gram \( c \) in the books published in year \( j \), which we denote \( x(c_j) \).

Since the raw prevalence \( x(c_j) \) will fluctuate with the total volume of text published in a given year \( j \), we denote \( V_j \), we normalize the prevalence \( x(c_j) / V_j \) with an estimate of \( V_j \). We use the total number of yearly occurrences of end-of-sentence punctuation to estimate the total volume of text published, \( N_j = X(\text{.}) + X(\text{!}) + X(?) \); hence, the volume-normalized time series of n-gram \( c \), is given by \( \hat{x}(c_j) = x(c_j)/N_j \) for every year \( j \). Note that \( N_j \) roughly corresponds to the total number of sentences published, because nearly all sentences are terminated by end-of-sentence punctuation.

However, the magnitude of yearly normalized prevalence values will differ significantly between n-grams of different lengths; e.g., “never” is likely much more prevalent than “they will not believe” for the same volume published, since the former is a 1-gram and the latter a more specific 4-gram. Nevertheless, both may follow a similarly shaped pattern of historical change. Furthermore, the volume of books published increases rapidly over time; hence the variance of the time series of n-gram occurrence may also change over time. To allow comparisons of the patterns of changing prevalence over time between series of different magnitudes and variance, we subtract the 1895 to 2019 mean from all prevalence time series results under 10,000 random variations of the set of CDS n-grams (e.g., by making different CDS n-gram choices) we calculate the 95% confidence intervals for this distribution by a 10,000-fold random resampling with replacement of the respective set of CDS n-grams (i.e., US English, Spanish, or German). Each such random resample results in a new mean time series for the given random set \( C \) of n-grams, i.e., \( z_{1885,2020}(c_j) \). The resulting yearly distribution of z scores indicates how much our yearly results can vary as a result of random changes to our CDS n-gram set and thereby tests the robustness of our time series results under 10,000 random variations of the set of CDS n-grams. All time series are normalized and converted to z-scores before resampling; hence we obtain a distribution of yearly z scores from which the relevant percentiles, including the median, are determined.

Data Availability. Previously published data were used for this work (https://storage.googleapis.com/books/ngrams/books/datasetsv3.html).

ACKNOWLEDGMENTS. J.B. is grateful for support from the Urban Mental Health Institute of the University of Amsterdam, Wageningen University and Research, the NSF (NSF Social, Behavioral and Economic Sciences [SBE 1636636], and the support of the Indiana University Vice Provost for COVID-19 Research. We thank Jonathan Haidt of New York University who provided the inspiration for this investigation by speculating on a recent societal shift toward a style of thinking that may be associated with internalizing disorders.

Bollen et al. 6 of 7 | PNAS

https://doi.org/10.1073/pnas.2102061118


