



**Centre for Advanced Research in Experimental and Applied Linguistics
(ARiEAL)**

Title: Concreteness and Psychological Distance in Natural Language Use

Journal: Psychological Science

Author(s): Snefjella, B., Kuperman, V.

Year: 2015

Version: Post-Print

Original Citation: Snefjella, B., & Kuperman, V. (2015). Concreteness and Psychological Distance in Natural Language Use. *Psychological Science*, 26(9), 1449–60. <https://doi.org/10.1177/0956797615591771>

Rights: © <2015>. This is the post-print version of the following article which was originally published by Psychological Science in 2015: Snefjella, B., & Kuperman, V. (2015). Concreteness and Psychological Distance in Natural Language Use. *Psychological Science*, 26(9), 1449–60. <https://doi.org/10.1177/0956797615591771>

If you would like to learn more about ARiEAL research centre, please visit us at:

W: ariefal.mcmaster.ca **T:** [@ARiEAL_Research](https://twitter.com/ARiEAL_Research)

Concreteness and Psychological Distance in Natural Language Use

Snefjella, B.*, Kuperman, V.

McMaster University

*Department of Linguistics and Languages, McMaster University, Togo Salmon Hall 626, 1280 Main St. West, Hamilton, Ontario, Canada L8S 4M2

Abstract

Existing evidence shows that more abstract mental representations are formed and more abstract language is used to characterize phenomena that are more distant from the self. Yet the precise form of the functional relationship between distance and linguistic abstractness is unknown. In four studies, we tested whether more abstract language is used in textual references to more geographically distant cities (Study 1), time points further into the past or future (Study 2), references to more socially distant people (Study 3), and references to a specific topic (Study 4). Using millions of linguistic productions from thousands of social-media users, we determined that linguistic concreteness is a curvilinear function of the logarithm of distance, and we discuss psychological underpinnings of the mathematical properties of this relationship. We also demonstrated that gradient curvilinear effects of geographic and temporal distance on concreteness are nearly identical, which suggests uniformity in representation of abstractness along multiple dimensions.

Keywords

psychological distance; construal-level theory; embodied cognition; social media; Twitter; abstraction; concreteness

Introduction

One of the fundamental and unique abilities of the human mind is to transcend the boundaries of here and now: to imagine distant times, far-away places, and other people. The psychological mechanism of abstraction underlies this mental ability, but how this mechanism operates is a matter of continuing debate (Barsalou, 2008; Boroditsky & Ramscar, 2002; Burgoon, Henderson, & Markman, 2013; Fischer & Zwaan, 2008; Gallese & Lakoff, 2005; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Paivio, 1990; Schwanenflugel, Harnishfeger, & Stowe, 1988). Yet social psychologists have noted a positive correlation between the perceived distance of an object or event and the level of abstraction at which that event is represented mentally. For instance, the influential construal-level theory of psychological distance (Trope & Liberman, 2003, 2010) states that objects and events that are proximal (close to an egocentric self) are represented with rich, complex, concrete, contextual, and subordinate-level features. This is referred to as a low-level construal. A high-level construal is the representation of distal objects and events abstractly by their simple, invariant, superordinate-level characteristics. For example, if we are preparing a lecture for tomorrow (a proximal event), we will worry about which room to go to. When preparing a lecture for next month (a distal event), we will worry about its topic. According to construal-level theory, the distance-driven differences in construal arise because abstract representations and goals are more stable over time than concrete representations (the topic of my lecture remains the same, even if the location changes). Thus, abstraction leads to successful traversing of psychological distances (Trope & Liberman, 2010).

The relationship between abstraction and psychological distance is implicated in many personal and social phenomena, including the consistency of attitudes and evaluations in an individual (Ledgerwood,

Trope, & Chaiken, 2010), the actor-observer bias (Nisbett, Caputo, Legant, & Marecek, 1973), moral judgments (Amit & Greene, 2012), politeness (Stephan, Liberman, & Trope, 2010), subjective judgments of truth (Hansen & Wanke, 2010), and consumer preferences (Fiedler, 2007). The hypothesized positive correlation between the abstractness of mental representations and psychological distance has received support in many experimental paradigms and measures, including action identification (Fujita, Henderson, Eng, Trope, & Liberman, 2006; Liviatan, Trope, & Liberman, 2008), a “distance Stroop” task (Bar-Anan, Liberman, Trope, & Algom, 2007), and surveys and questionnaires (Eyal, Liberman, Trope, & Walther, 2004; Trope & Liberman, 2000; Wakslak, Trope, Liberman, & Alony, 2006; for a recent meta-analysis of research on the construal-level theory, see Soderberg, Callahan, Kochersberger, Amit, & Ledgerwood, 2015). Of greater importance for the present research are studies that capitalized on the ability of language to reflect abstractness of mental representations through abstractness of expressed meanings (Paivio, 1990; Schwanenflugel et al., 1988). These tasks elicited linguistic productions from participants while manipulating the distance of what participants were prompted to write about. Psychological distance of described phenomena, typically conceptualized as their construal level, was then operationalized as relative abstractness of produced texts. A robust finding across the studies was that more abstract language is used to characterize phenomena that are more distant from the self temporally, spatially, or socially or are more hypothetical (Trope & Liberman, 2010). Although valuable, the current methodology of using linguistic productions to study the link between abstraction and psychological distance is limited. In a typical laboratory experiment, small groups of undergraduate participants are prompted to write about distant or near objects or events, which often results in small samples of language obtained from

individuals with a relatively homogenous age range and experience. The scale of data collection is further limited by labor-intensive manual coding of linguistic abstractness. For instance, Fujita et al. (2006), Stephan et al. (2010), and Gong and Medin (2012) coded the writing of participants using the linguistic-categorization model of Semin and Fiedler (1988). In a similar vein, Alter and Oppenheimer (2008) used three human coders to rate productions of participants as being abstract, concrete, or both.

Even more drastically, instead of being treated as a continuum, abstractness is routinely binned into discrete categories. As the meta-analysis of research on construal-level theory further shows, distance was dichotomized into “close” and “far” categories in all 267 studies that met the authors’ inclusion criteria (Soderberg et al., 2015). Either categorization obscures the precise mathematical form of their functional relationship and prevents characterization of the effect of abstractness on psychological distance in a graded way.

In the present study, we addressed these drawbacks by using norming megastudies with ratings of semantic word properties as well as vast collections of language productions in text corpora. We used these resources to examine—on a larger scale and with a broader range of examples than in previous studies—abstractness of linguistic productions that describe objects or events positioned at various psychological distances. In our study, operationalization of construal level relied on a recent data set of concreteness ratings of 40,000 English words (Brysbaert, Warriner, & Kuperman, 2014). Ratings of words were made on a scale from 1 (abstract) to 5 (concrete) and averaged over 30 participants: Resulting concreteness norms ranged from 1.04 (“essentialness”) to 5 (“pitbull”). Using this data set, we were able to measure abstractness in language without human coders and with any number of productions.

By handing the drudgery of coding to a computer, we were able to study

psychological distance at scales not previously possible: millions of observations from thousands of language speakers varying in age, socioeconomic status, personality traits, and place of residence. Furthermore, we analyzed natural language use, not language elicited in an experimental task; our study tested the ecological validity of the construal-level theory and complements experimental research. The meta-analysis by Soderberg et al. (2015) pointed to the inability of current studies of psychological distance to identify the precise mathematical form of the gradient functional relationship between distance and abstractness. Using clever analytical techniques, Soderberg et al. (2015) predicted the relationship to be curvilinear. Our approach charted the relationship along a continuum of psychological distances and provided other psychologically meaningful interpretations of its mathematical properties.

We present four studies, each of which examined one of the critical dimensions of construal-level theory; all were based on social-media sources. Specifically, we tested whether more abstract language is used in textual references to more geographically distant cities (Study 1), time points further into the past or future (Study 2), and references to more socially distant people (Study 3). In Study 4, we examined whether a theme that is commonly experienced across all times and distances—death and dying—follows the pattern observed in the aggregate data.

Study 1: Geographic Distance

Method

This study explored the role of geographic distance in explaining variability in the level of construal of U.S. cities. Construal was operationalized as concreteness of language used in relation to the cities. We selected the 30 most populous U.S. cities (with 600,000 inhabitants as an arbitrary lower threshold) using rankings of their population size in 2013 (http://en.wikipedia.org/wiki/List_of_United_States_cities_by_population),

with the exception of Washington, D.C., as the city name is homonymous with a name of a geographically distant state. New York City and Oklahoma City, while homonymous with their respective states, are embedded in those states and thus were expected to introduce less noise in distance estimates.

Social media—a source of millions of data points, a broad selection of geographic objects, and a full range of possible distances—enables an expansion and refinement of prior experimental studies. Using the publicly available data stream from the Twitter application program interface at <https://dev.twitter.com>, we collected tweets that were (a) geo-tagged (i.e., had geographic information system coordinates of the location where the tweet was produced), (b) were sent from within the United States, and (c) contained the name of one of the major U.S. cities included in our analysis. To calculate the geographical distance between the location of the tweet production and the city of reference, we used the latitude and longitude coordinates of that tweet and of that city's center, as supplied by the `geocode()` function of the `ggmap` package (Kahle & Wickham, 2013) in R statistical software (Version 3.01; R Development Core Team, 2013). We applied the Haversine formula to obtain the greatcircle distance between these two points. To remove distributional skewness, we log-transformed (base 10) all distances. All calculations and analyses in this and subsequent studies were made using R software. It is possible that a more psychologically valid measure of geographic distance is the time it takes to commute from the tweet location to the respective city center. We opted for the great-circle distance because estimates of the commute time (including waiting time in airports and traffic hours) are inherently variable.

Results

A total of 712,198 tweets satisfying our criteria were collected between March and May 2014. Twitter is trendy and

dynamic; collecting tweets over a wide time frame helps prevent trending topics from influencing the results. A further trimming retained tweets that contained four or more words (excluding the city name) that had concreteness ratings in Brysbaert et al.'s (2014) data set: An average tweet contained 8.42 such words (SD = 3.9). This reduced the pool to 478,920 geo-tagged tweets. We calculated the mean concreteness of each tweet on the basis of words with available ratings (M = 2.83, Mdn = 2.78, range = 1.46– 4.92, SD = 0.47). Thus, the degree of construal was operationalized as a prevalence in the tweet of words that were rated in Brysbaert et al.'s (2014) study as more concrete or more abstract when presented out of context.

Geographic distances between the point of origin of tweets and cities they referred to ranged from 0.1 km to 4,220.9 km (M = 516.7, Mdn = 24.3, SD = 885.7; see Fig. 1 for the distribution of distances). To decrease noise in the raw data, we binned observations into percentiles of the log distance distribution and calculated mean concreteness of each bin. A cubic function ($y = 3.05 + 0.03x - 0.18x^2 + 0.04x^3$) provided an optimal polynomial fit to concreteness in a linear regression model with raw polynomials of log distance as predictors (adjusted R² = .88; see Fig. 1). Table 1 reports the goodness of fit of the cubic function as well as lower- and higher-order polynomial functions using hierarchical regression: Successive models were compared using the `anova()` function in R. While a cubic function was the best polynomial fit to the data, other functional relationships might offer a similarly good fit (e.g., logistic curve). Because there is no theoretically grounded expectation as to the form of the curvilinear dependency, we leave exploration of alternative functional forms to future work.

The pattern in Figure 1 indicates a substantial decrease in the concreteness of tweets regarding cities with an increase in the distance of the tweeting person from that city: The drop in concreteness between the

extreme distances was estimated at 0.5 points of concreteness or 12.5% of the concreteness scale. This pattern, based on nearly half a million observation points and 30 cities, is perfectly in line with the predictions and experimental validations of the positive correlation between psychological and geographic distance. Moreover, we confirmed the predictions of Soderberg et al. (2015) that construal and distance have a curvilinear relation.

The functional form of the fitted curve, and its excellent fit to the data, enabled us to further interpret parameters of the cubic function. The inflection point (i.e., the point at which the second derivative of the function changes its sign) was estimated at 1.5 (in log-10 units) or around 30 km from the city center. This implies that as a typical tweeting person moves away from the city center to this point, the concreteness of his or her references to the city decreases at a relatively high rate. This decrease becomes less precipitous as distances increase beyond 30 km, and, as the first derivative of the polynomial shows, there is little to no decrease in concreteness associated with distances above 100 km.

We further speculated that the inflection point is psychologically meaningful and demarcates a distinction between (a) being within a city, where concreteness of mental representation of that city is the highest (the level of construal the lowest) and decreases sharply from the city center to the outskirts and (b) being outside of the city, where the representation of the city is more abstract overall (the level of construal is higher) and is less affected by how far the Twitter user is from the city. To test this hypothesis of immediacy of experience, we calculated the radius of the city area for each of the 30 cities, under a simplifying assumption that cities have a perfect circular shape (estimates of the city area were obtained from <http://www.citymayors.com/statistics/largest-citiesarea-250.html>). Extreme radii were found for New York, New York ($r = 53$ km) and El Paso, Texas ($r = 13$ km), and the mean

radius was 25 km, close to the inflection point of the fitted curve at 30 km. While more sophisticated measurements of the urban territory will be necessary, the observed value is consistent with the notion that the construal level increases (and concreteness of language decreases) more drastically as the speaker loses the immediacy of the urban experience when moving from the city center to its outskirts: Once outside the city, the construal level is more stable and high.

Study 2: Temporal Distance

Method

A robust finding in the literature on memory and prediction in relation to psychological distance is that remoter events, whether in the past or the future, elicit a higher level of construal (Trope & Liberman, 2000, 2003). To test this dimension of psychological distance, we employed the Usenet corpus collected by Shaoul and Westbury (2013), which consists of over 7 billion word tokens of public Usenet postings collected from 47,860 Englishlanguage news groups between October 2005 and January 2011. Several temporal terms were used to examine effects of temporal distance on concreteness of language in which past and future events are described. We explored distance both within specific time units (e.g., 10 years ago vs. 100 years ago) and between units (e.g.,_days from now vs. centuries from now; last week vs. next week).

Study 2a: “years ago.” Soderberg et al. (2015) predicted the curvilinear relationship between distance and construal on the basis of their meta-analysis, which placed different studies of temporal distance along an objective timeline from 1 to 365 days. We were unable to re-create this finding in the corpus, as Usenet contributors almost never refer to distances over 10 days with the phrase “X days ago.” However, phrases such as “X years ago” yielded intriguing results.

We identified all occurrences in the corpus of the phrase “X years ago.” We

further extracted 5 words to the left and 5 to the right of the critical phrase, “years ago” (e.g., “wind generation of electricity 30 years ago and they were commonplace then”). The 10-word window around the target word (henceforth referred to as the context) was chosen to approximately equate the number of words in the Twitter (with its 140-character limit per tweet) and Usenet extracts. We leave for further study the question of which window size provides the best accuracy. Numbers preceding the target phrase (written either as words or numerals) served as the metric of the temporal distance from the time of e-mail submission to Usenet. We restricted the time range to 1 through 999 years ago. Finally, we removed all extracts with contexts that contained fewer than 4 words with concreteness ratings available in Brysbaert et al.’s (2014) data set.

Study 2b: “ago” versus “from now.” We were also interested in comparing the abstractness with which people refer to distances in the past or future, as measured by different time units. In the corpus, we identified all occurrences of the phrases “X ago” and “X from now” in which X was a unit of time: minute, hour, day, week, month, year, decade, or century. We further extracted five words to the left and five to the right of the critical phrase (e.g., “centuries ago,” as in “situation as recent as two centuries ago when much academic instruction was”). Numbers (written either as words or numerals) were removed from the preceding context window. The resulting scale of time units was then ordinal (a week ago is further in the past than a day ago), rather than continuous, as in Study 2a. Finally, we removed all extracts with contexts that contained fewer than four words with concreteness ratings in Brysbaert et al.’s (2014) data set.

Study 2c: “last” versus “next.” To ensure that observed differences between time units were not an artifact of our choice of the language denoting temporal distance (time units from now and time units ago), we conducted an additional set of analyses using contexts for phrases such as “yesterday,”

“tomorrow,” “last week,” “next week,” and “last month.” Contexts were defined as in Studies 2a and 2b, and the trimming procedures were the same as in Study 2b.

Results

Study 2a: “years ago.” A total of 265,859 extracts containing the phrase “years ago” were identified in the Usenet corpus. Because of skewness, temporal distances in years were log-transformed (base 10). Observations were binned by their temporal distance into 36 intervals with open left boundaries—formed by the numbers 1 to 19 (in increments of 1), 20 to 90 (in increments of 10), and 100 to 1,000 (in increments of 100)—and closed right boundaries. The histogram of the distribution of temporal distances is shown in Figure 2a. Mean concreteness of contexts was calculated for each bin and plotted against the log 10 of the numeral in the interval’s left boundary (see Fig. 2a).

As with geographic distance, the best polynomial fit to concreteness was obtained with a cubic curve ($y = 2.44 + 0.11x - 0.12x^2 + 0.03x^3$) and showed an excellent fit in a linear regression model with raw polynomials of log temporal distance (adjusted $R^2 = .80$; see Table 1 for model comparison). Verbal descriptions of past events were the more abstract (i.e., construed at a higher level) the more years had passed since the described event. The drop in concreteness between the extremes of the temporal range was fairly small and amounted to approximately 0.1 units of concreteness or 2.5% of the concreteness scale. Again, the observed pattern converged with the experimental evidence of the construal-level theory of psychological distance, in which construal is operationalized as concreteness of verbal description of the event. This also shows that the predicted curvilinear relationship (Soderberg et al., 2015) between construal and distance holds for multiple dimensions of psychological distance.

A further inspection of the functional curve pointed to a faster drop in concreteness in verbal representations of events up to the inflection point of 1.57 log units (or 37 years ago), a slower decrease in concreteness for more distant events past the inflection point, and virtually no change in concreteness of contexts associated with events taking place 200 to 1,000 years ago. Drawing an analogy to geographic distance, we speculated that the inflection point demarcated a change in the immediacy of one's experience with events that happened during one's lifetime and those that preceded it. As the literature on collective and generational memory demonstrates, critical social events (wars, natural disasters, changes in political regime) are more salient in mental representations of the past in those individuals who were exposed to the event as it happened than in individuals who were not (Pennebaker, Paez, & Rim, 2013). If true, we would expect the inflection point of the functional curve (37 years of age) to be close to the find age data on Usenet users, the available statistics on Internet users did not diverge from this number (e.g., the average age of social-media users in 2012 was 37.9; Pingdom, 2014; see also Eisenstein, in press). Again, the functional form of the effect of temporal distance on concreteness of language production suggests immediacy of one's experience with events as an important factor in the construal level of mental representation and linguistic expression of those events.

Study 2b: "ago" versus "from now." A total of 767,842 critical phrases and their surrounding contexts were identified with time units (minutes to centuries) followed by "ago" or "from now." After trimming, the data pool contained 698,391 contexts. Mean concreteness was calculated for each context and plotted against respective time units. Figure 2b summarizes the functional relationship between temporal distance from "now" (the time of writing of the posting) and the level of construal of the temporally marked event.

The data in Figure 2b indicated a near-linear decrease in concreteness for events that are further away from the present on the ordinal scale of time units, in accordance with the hypothesized relationship between temporal and psychological distance. The maximum contrast between time units (hours ago and centuries ago) was 0.2 units of concreteness, corresponding to 5% of the concreteness scale. Regression models further indicated large effect sizes (R^2 s = .89 and .55 for past and future, respectively). The patterns also showed that there was a preference for talking about past rather than future events, as evidenced in the circle sizes, proportional to log frequency of the phrase occurrence in the corpus. The intercepts of the regression lines further suggested that overall past experiences are represented in more detail (higher concreteness) than events that are envisioned in the future, which is in line with experimental research into the mental simulation of past and future events (D'Argembeau & Van der Linden, 2004; Johnson, Foley, Suengas, & Raye, 1988).

Study 2c: "last" versus "next." There were a total of 1,025,121 contexts for phrases such as "last month" versus "next month." Figure 2c summarizes the results for all time units. We again noted a decrease in concreteness as temporal distance from the present increased. However, with these key phrases, the past and future appeared to be (almost) mirror images of each other, similar both in the slopes of the regression lines ($\beta = 0.08$ for the past and $\beta = -0.08$ for the future events), the amount of explained variance (R^2 s = .89 and .91, respectively), and (log-10) frequencies of occurrence of respective phrases, shown as circle sizes in Figure 2c. Also, the contrast between maximally different time units ("yesterday/ today" and "last/next century") was much larger than in the comparison in Study 2b ("days ago/from now" vs. "centuries ago/from now") and amounted to 0.4 units of concreteness or 10% of the concreteness scale.

Study 3: Social Distance

Method

Using the same procedure as in Study 2, we next investigated social distance by extracting from the Usenet corpus five words on each side of a target word. Previous experiments (Liviatan et al., 2008; Stephan et al., 2010) have shown that psychological distance between individuals is perceived to be larger, and the level of construal higher, if a social relationship between those individuals is more distant. To operationalize closeness of social relationships in a corpus, we took as a point of departure the Bogardus Social Distance Scale (Bogardus, 1933; see also Parrillo & Donoghue, 2005). The scale evaluates the degree of willingness to establish social contacts with representatives of a racial, ethnic, socioeconomic, occupational, or other social group. The scale identifies closeness as the individual's willingness to accept the group representatives using a 7-point scale: (a) potential partners in marriage, (b) close friends, (c) neighbors on the same street, (d) coworkers in the same occupation, (e) citizens in the same country, (f) only visitors to his or her country, or (g) people to be excluded from his or her country. To adapt the scale to the observed data, we converted the scale from a cumulative one (i.e., agreement with a higher degree of closeness implies agreement with all lower-degree categories) to a discrete ordinal one by identifying terms belonging to each of the scale's categories (e.g., "friend," "ally," "confidant," "pal," "chum," "buddy" for the close-friends category and "compatriot," "countryman," "countrywoman" for the visitors category. The full list of 39 terms—created using our linguistic intuitions and the Merriam-Webster thesaurus (<http://www.merriam-webster.com/>)—is reported in Table 2.

Results

After trimming, 422,553 data points remained. Figure 3 plots the mean

concreteness of the contexts, grouped by social-distance categories, against the ordinal scale of social distance. Although the number of data points and confidence intervals varied across categories, the overall trend was in agreement with the hypothesized link. Groups of individuals that are considered more distant socially are also construed in less concrete terms, with a maximum contrast of 0.15 points (about 4% of the available concreteness scale) between the family members and foreigners categories.

Study 4: Concreteness of the Theme of Death Over Time and Geographic Distance

Method

Two points of criticism can be raised with regard to Studies 1 through 3. First, we used aggregate measures of the concreteness of verbal contexts, which average over a multitude of phenomena and a diversity of personal and collective experiences, and thus might lead to ecological fallacy (Robinson, 1950). Second, there are alternative explanations as to why a person might choose more abstract over more concrete words when describing a remote phenomenon. It might be due to the linguistically faithful reflection of the distance-driven change in one's mental representation, which would be consistent with the premises of construal-level theory. Conversely, it might take place because one does not have direct experience with the phenomenon and has access only to its gistlike representation through language; thus, one can describe it only in abstract terms. This relative abstractness is not expected to vary with distance but only with the amount of experience.¹ In Study 4, we considered construal of events related to death as a function of geographic and temporal distance. Death is a concept that is acquired early, is salient and memorable as an event, and occurs to all living beings at all times and all locations, which gives on average an equal probability to directly or indirectly experience (somebody else's) death at all

distances from the self. Finding a curvilinear relationship between the concreteness of texts constrained to a familiar, ubiquitous event and distance from the self would be a step toward ensuring that the aggregate patterns are made of converging individual patterns and that—at least in some cases—predictions of construal-level theory are due to the change in distance and not only to the change in the strength of personal experience.

To address these issues, we extracted 354 Twitter messages containing the words “died,” “dead,” or “death” from the data pool of Study 1 and 3,735 contexts from Usenet e-mails containing the same words and a target phrase, “years ago,” as in Study 2a. Mean concreteness of those tweets and those contexts were calculated for each bin of log geographic distance (in km) and log temporal distance (in years), with bins defined as in Study 1 and Study 2a, respectively. A similar study of social distance was not feasible because some of the categories (e.g., compatriots) did not offer a sufficient sample size to allow comparison.

Results

Concreteness and log geographic distance of texts related to death demonstrated a curvilinear relationship, which was well approximated by a cubic polynomial function ($y = 2.96 - 0.05x - 0.1x^2 + 0.03x^3$; $R^2 = .18$). The top panel of Figure 4 both reports the relationship for the tweets referencing death (solid line) and—for reference—replicates at a different scale the curve from Figure 1 (dotted line) that summarizes the trend in all tweets about major U.S. cities in Study 1. The thematically constrained subset of tweets showed a similar if slightly flatter pattern than the overall trend. Tweets about death sent from the city center were maximally concrete; their concreteness dropped dramatically when outside of the city and leveled off at distances above 100 km, with a slight increase in concreteness at very remote distances. Similarly, the concreteness of Usenet contexts

containing the words “years ago” and death-related words was a sigmoid function of log temporal distance, which was well approximated by a cubic polynomial ($y = 2.53 + 0.41x - 0.40x^2 + 0.08x^3$; $R^2 = .48$). The bottom panel of Figure 4 plots the curve estimated for death-related messages (solid line) and the overall trend for all messages (dotted line), which replicates, with a correction for scale, the curve in Figure 2a. Death-related contexts were generally more concrete than the thematically unconstrained contexts, but much like the overall trend in Study 2a (Fig. 2a), they showed the maximum of concreteness for deaths that occurred very recently, a drastic decrease in concreteness as the past became less recent, and a leveled-off pattern after some three decades from the time the message was written. To sum up, the curvilinear relationship between language concreteness and log (geographic and temporal) distance was confirmed even with a constraint that focused on one class of phenomena (i.e., those related to death). Thus, phenomena that are likely to be part of individual experience and have a similar probability of occurring in an individual life recently or a long time ago, close or far, are construed with a similar level of detail at different distances as the entirety of phenomena that our method captures in a language corpus.

General Discussion

We present a new method of examining an aspect of embodied cognition (Barsalou, 2008; Fischer & Zwaan, 2008; Gallese & Lakoff, 2005; Meteyard et al., 2012): the interplay among perceived and objective distance, abstraction as a mental faculty, and abstractness as a property of language. We identified words or phrases that denoted an entity or event for which we had information about distance. This information could be encoded in the phrase (“spouse” vs. “coworker” for social distance) or explicitly stated as a number (“twenty

years from now”). We then measured the concreteness of language that co-occurred with that word or phrase in texts and correlated this concreteness with distance. The utility of our method was demonstrated in studies of three critical dimensions of construal-level theory: spatial, temporal, and social distance. In all four studies, the predictions of construal-level theory held. Tweets containing the names of an American city become more abstract as the geographic distance between the person sending the tweet and that city increased. Similarly, verbal contexts of time points further into the past or future tended to be more abstract, as did verbal descriptions of more socially distant people.

Our use of multiple linguistic expressions of distance and massive amounts of linguistic productions in corpora allowed us to go beyond validation of prior experimental findings and answer outstanding questions about construal-level theory. One theoretical point raised by Soderberg et al. (2015) was whether distance increased the processing of abstract information, decreased the processing of concrete information, or both. We note that in all our studies, greater distance led to more overall abstractness in language, but at every distance, the linguistic productions showed gradience in their abstractness and concreteness. Specifically, our regression analyses of two continuous metrics of spatial (kilometers from the city) and temporal (years before writing) distance revealed that the relationship between log distance and abstractness of language is curvilinear and is well approximated by a cubic polynomial curve. Language used in relation to cities and events is at its most concrete (construal is at its lowest) when the experience of that city or that event is most immediate (e.g., being in the city center or occurring in the very recent past; cf. Hirst et al., 2015). Tweets become abstract more rapidly between a city center and its suburbs than between the city boundary and any other location in the country. Time references become abstract more rapidly

between the present and the time point in the past that indicates a typical life span than between events in the distant and very distant past. This is also true when texts were thematically constrained to refer to death-related phenomena. Thus, we both confirmed and specified the curvilinear relationship between distance and abstraction predicted by Soderberg et al. (2015). Moreover, the similarity of effects that physical and temporal distance have on linguistic concreteness—displayed over all relevant contexts or only a thematically constrained set—corroborates the long-standing observation that language often expresses temporal relations via metaphors of space (Boroditsky, 2000, 2001; Boroditsky & Ramscar, 2002). Finally, symmetrical effects of past and future temporal distance on concreteness suggest analogous cognitive processes involved in remembering past events and imagining future ones (e.g., Schacter, Addis, & Buckner, 2007).

As with any method, a corpus-based approach has limitations. We were unable to explore the construal of entities that did not correspond to a word or phrase or that did not occur in corpora with sufficient frequency. There is little doubt that noise was introduced into the data from homography and polysemy: “bank” as a financial institution and the edge of a river, “work” as a noun and a verb, “Chicago” as a city and a musical. Also, we used concreteness ratings for words presented out of context to calculate the average concreteness of sequences of words that occur in context, missing out on metaphoric word use and other context-driven changes in word meaning. It is improbable, however, that our patterns arose because of a systematic bias in our operationalization of context concreteness, as this would have required contexts not only to be consistent in how they changed word concreteness but also to modulate this amount and direction of change as a function of distance. Many of these limitations are addressable: By carefully restricting the searched linguistic materials and their

contexts (exemplified partly in Study 4); restricting the age, gender, or place of residence of contributors (as self-reported in several social-media sources); or taking temporal cross-sectional snapshots of the data. The utility of an observational approach based on corpora as a complement to experimental studies outweighs its limitations. It has the advantages of (a) ecological validity through observation of psychological distance in texts produced in natural communicative settings; (b) automatized ability to track psychological distance in vast spans of language created by heterogeneous, large populations; and (c) ability to investigate a very broad or a very focused range of entities or events. For instance, we chose American cities as our geographical objects of interest. Any object for which we have a name and latitude and longitude coordinates can be explored for the effect of geographic distance on construal with the method presented in this article. Equally, choosing one theme or a specific time slice in a corpus enables one to break down the aggregate trends demonstrated here into any level of granularity. Notably, corpora and social media free researchers to study psychological distance outside of the laboratory.

Author Contributions

B. Snefjella developed the study concept. Both authors collected, analyzed, and interpreted the data. B. Snefjella drafted the manuscript, and V. Kuperman provided revisions. Both authors approved the final version of the manuscript for submission.

Acknowledgments

We thank the Sherman Centre for Digital Scholarship and the Research & High-Performance Computing Support group at McMaster University for technical support. We also thank Emmanuel Keuleers and two

anonymous reviewers, as well as the audience of the 55th Annual Meeting of the Psychonomic Society and the 2015 Annual Meeting of the American Association for the Advancement of Science for providing valuable feedback.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by Social Sciences and Humanities Research Council Insight Development Grant No. 430-2012-0488, Natural Sciences and Engineering Research Council of Canada Discovery Grant No. 402395-2012, National Institutes of Health Grant No. R01 HD 073288 (principal investigator: Julie A. Van Dyke), and an Early Researcher Award from the Ontario Research Fund to the second author.

Note

1. We are indebted to Emmanuel Keuleers for raising this point.

References

- Alter, A. L., & Oppenheimer, D. M. (2008). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychological Science, 19*, 161–167.
- Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science, 23*, 861–868.
- Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: Evidence from a Stroop task. *Journal of Experimental Psychology: General, 136*, 610–622.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.
- Bogardus, E. S. (1933). A social distance scale.

- Sociology & Social Research*, 17, 265–271.
- Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75, 1–28.
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1–22.
- Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13, 185–189.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904–911.
- Burgoon, E. M., Henderson, M. D., & Markman, A. B. (2013). There are many ways to see the forest for the trees: A tour guide for abstraction. *Perspectives on Psychological Science*, 8, 501–520.
- Central Intelligence Agency. (2014). *The World Factbook: North America: United States*. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/geos/us.html>
- D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, 13, 844–858.
- Eisenstein, J. (in press). Written dialect variation in online social media. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *Handbook of dialectology*. New York, NY: Wiley.
- Eyal, T., Liberman, N., Trope, Y., & Walther, E. (2004). The pros and cons of temporally near and distant action. *Journal of Personality and Social Psychology*, 86, 781–795.
- Fiedler, K. (2007). Construal level theory as an integrative framework for behavioral decision-making research and consumer psychology. *Journal of Consumer Psychology*, 17, 101–106.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61, 825–850.
- Fujita, K., Henderson, M. D., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science*, 17, 278–282.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22, 455–479.
- Gong, H., & Medin, D. L. (2012). Construal levels and moral judgment: Some complications. *Judgment and Decision Making*, 7, 628–638.
- Hansen, J., & Wanke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality and Social Psychology Bulletin*, 36, 1576–1588.
- Hirst, W., Phelps, E. A., Meksin, R., Vaidya, C. J., Johnson, M. K., Mitchell, K. J., . . . Olsson, A. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events. *Journal of Experimental Psychology: General*, 144, 604–623.
- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, 117, 371–376.
- Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1), 144–161. Retrieved from <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- Ledgerwood, A., Trope, Y., & Chaiken, S. (2010). Flexibility now, consistency later: Psychological distance and construal shape evaluative responding. *Journal of Personality and Social Psychology*, 99, 32–51.
- Liviatan, I., Trope, Y., & Liberman, N. (2008). Interpersonal similarity as a social distance dimension: Implications for perception of others' actions. *Journal of Experimental Social Psychology*, 44, 1256–1269.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48, 788–804.
- Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 27, 154–164.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.
- Parrillo, V. N., & Donoghue, C. (2005). Updating the Bogardus social distance studies: A new national survey. *The Social Science Journal*, 42, 257–271.
- Pennebaker, J. W., Paez, D., & Rim, B. (2013). *Collective memory of political events: Social psychological perspectives*. New York, NY:

- Psychology Press.
- Pingdom. (2014). *Report: Social network demographics in 2012*. Retrieved from <http://royal.pingdom.com/2012/08/21/report-social-network-demographics-in-2012/>
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8*, 657–661.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*, 499–520.
- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 558–568.
- Shaoul, C., & Westbury, C. (2013). *A reduced redundancy USENET corpus (2005–2011)*. Edmonton, Canada: University of Alberta. Retrieved from [http://www.psych.ualberta.ca/~westburylab/downloads/usenetcorpus.download.html](http://www.psych.ualberta.ca/~westbury/lab/downloads/usenetcorpus.download.html)
- Soderberg, C. K., Callahan, S. P., Kochersberger, A. O., Amit, E., & Ledgerwood, A. (2015). The effects of psychological distance on abstraction: Two meta-analyses. *Psychological Bulletin, 141*, 525–548.
- Stephan, E., Liberman, N., & Trope, Y. (2010). Politeness and psychological distance: A construal level perspective. *Journal of Personality and Social Psychology, 98*, 268–280. Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*, 876–889.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*, 403–421.
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*, 440–463.
- Wakslak, C. J., Trope, Y., Liberman, N., & Alony, R. (2006). Seeing the forest when entry is unlikely: Probability and the mental representation of events. *Journal of Experimental Psychology: General, 135*, 641–653.

Table 1
Results of Hierarchical Regressions Comparing Models Predicting Context Concreteness

Polynomial degree	Geographic distance (Study 1)			Temporal distance (Study 2a)		
	R2	Comparison with previous model		R2	Comparison with previous model	
		$\Delta R2$	p		$\Delta R2$	p
Linear	0.775	-	-	0.745	-	-
Quadratic	0.776	0.001	>0.5	0.749	0.004	0.47
Cubic	0.883	0.107	<0.001	0.814	0.65	<0.001
Quartic	0.885	0.002	>.5	0.819	0.005	0.4

Note: In Study 1, the predictor was log geographic distance from the city; in Study 2a, the predictor was log temporal distance from the event in the past. The cubic polynomial provided the best fit.

Table 2
Terms Used for the Social-Distance Groups Defined by Bogardus (1933)

Family	Friends	Neighbors	Coworkers	Compatriots	Visitors	Foreigners
husband	friend	Neighbor	coworker	compatriot	visitor	immigrant
wife	ally	neighbour	co-worker	countryman	tourist	foreigner
spouse	confidant	peer	colleague	countrywoman	traveler	outsider
consort	confidante	homie	collaborator			stranger
	alter ego	homeboy	workmate			emigrant
	second self	homegirl				nonmember
	pal					noncitizen
	chum					newcomer
	buddy					alien

Note: From left to right, the groups are arranged from the most proximal to the most distal.

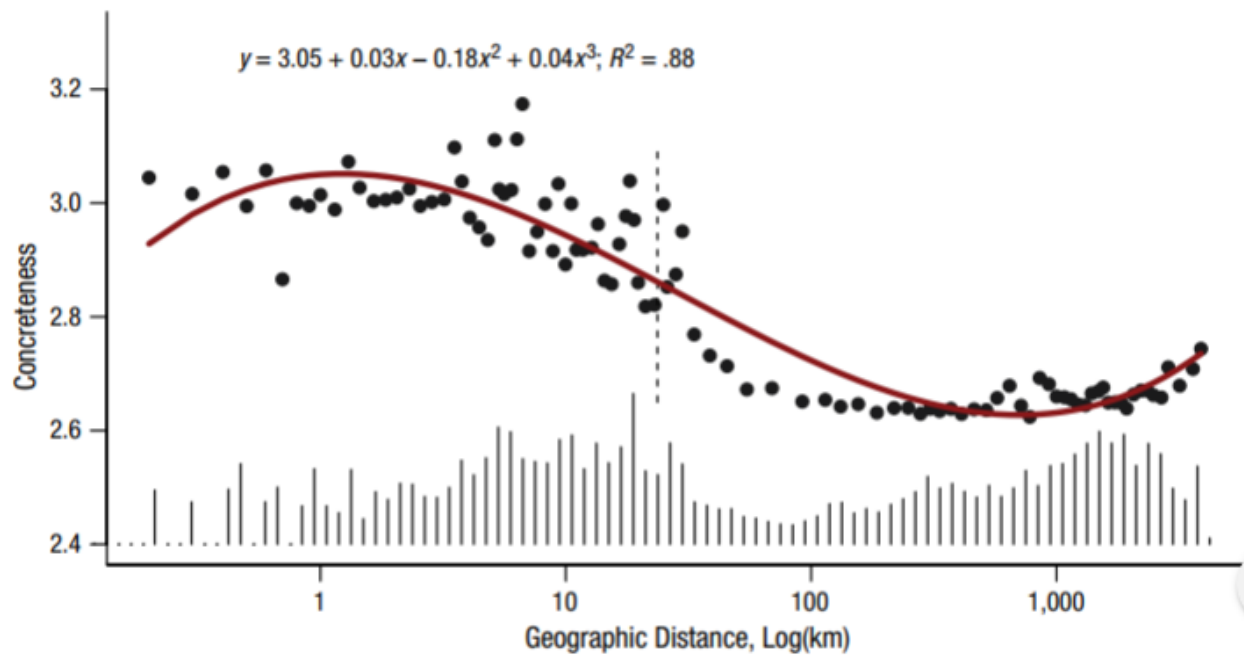


Fig. 1. Results from Study 1: scatterplot showing the mean concreteness of Twitter messages regarding U.S. cities as a function of log geographic distance. The dotted line represents the inflection point (i.e., the point at which the second derivative of the function changes its sign), and the histogram of distances is presented along the x-axis. The best-fitting regression line and equation are also shown.

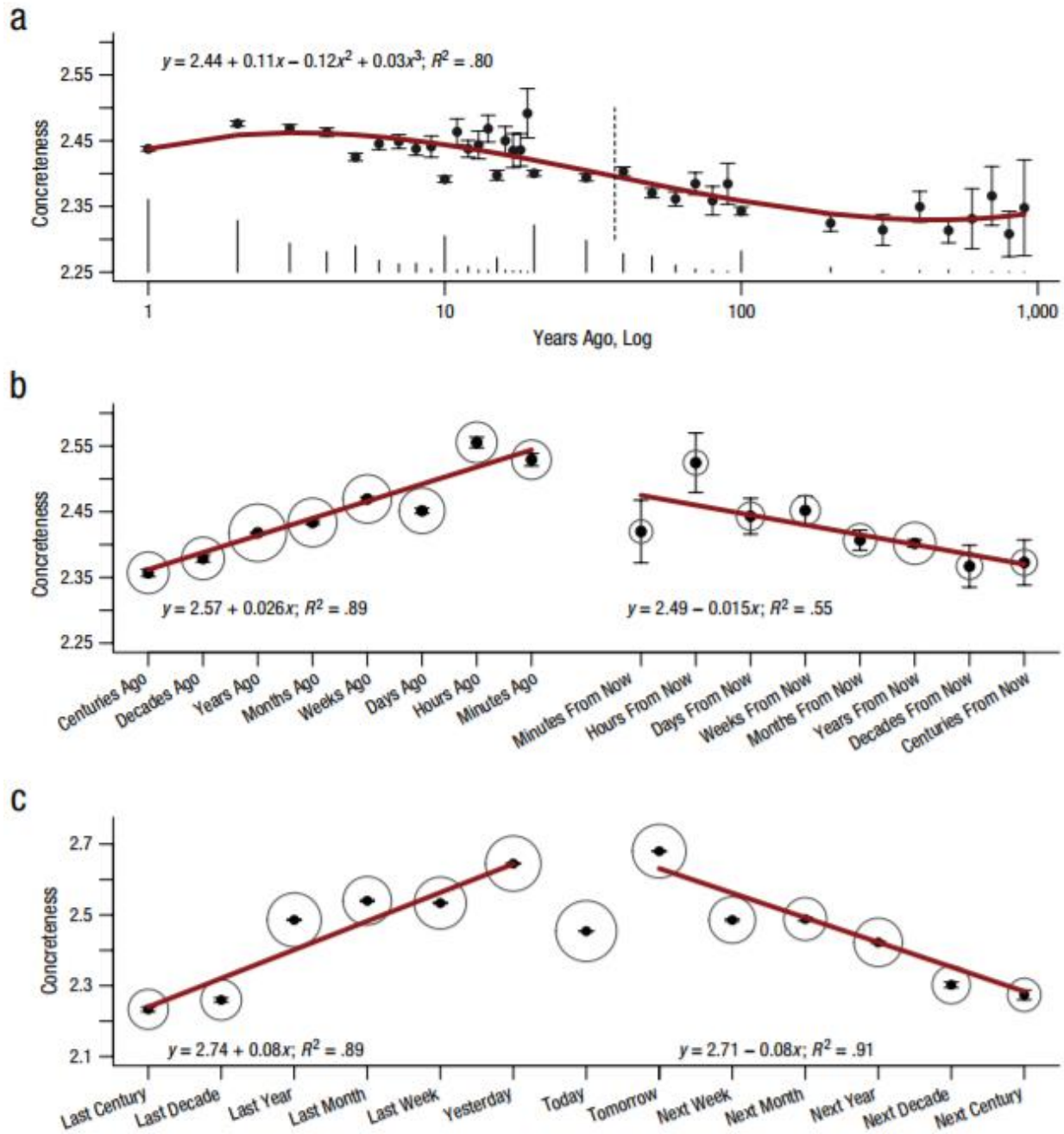


Fig. 2. Results from Study 2: mean concreteness of Usenet postings as a function of (a) log temporal distance from 1 to 999 years (Study 2a), (b) time units in the past (target phrase: “ago”) and the future (target phrase: “from now”; Study 2b), and (c) ordered time units in the past (target phrase: “last”) and the future (target phrase: “next”; Study 2c). In (a), the dotted line represents the inflection point (i.e., the point at which the second derivative of the function changes its sign), and the histogram of distances is presented along the x-axis. In (b) and (c), circle size is proportional to log-10 frequency of the target phrase. In all panels, the best-fitting regression line and equation are provided, and error bars (where visible) reflect 95% confidence intervals

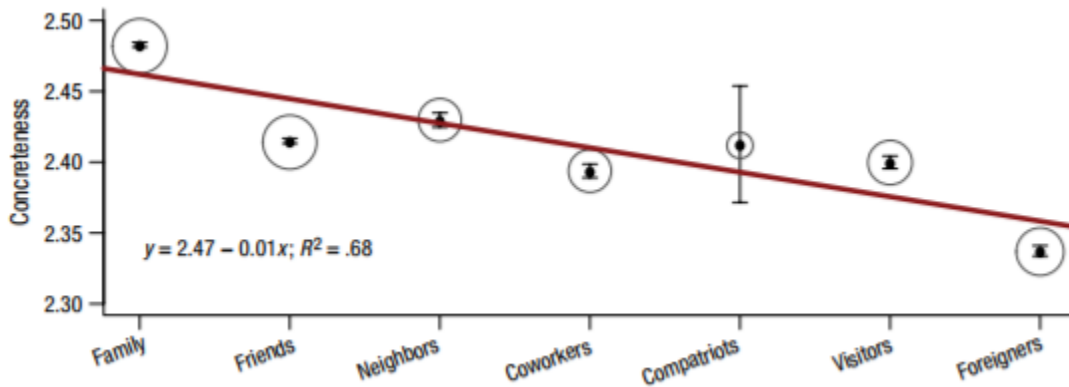


Fig. 3. Results from Study 3: mean concreteness of Usenet postings as a function of social-distance group (defined by Bogardus, 1933). Circle size is proportional to the log-10 frequency of search terms in each category. The best-fitting regression line and equation are provided, and error bars reflect 95% confidence intervals.

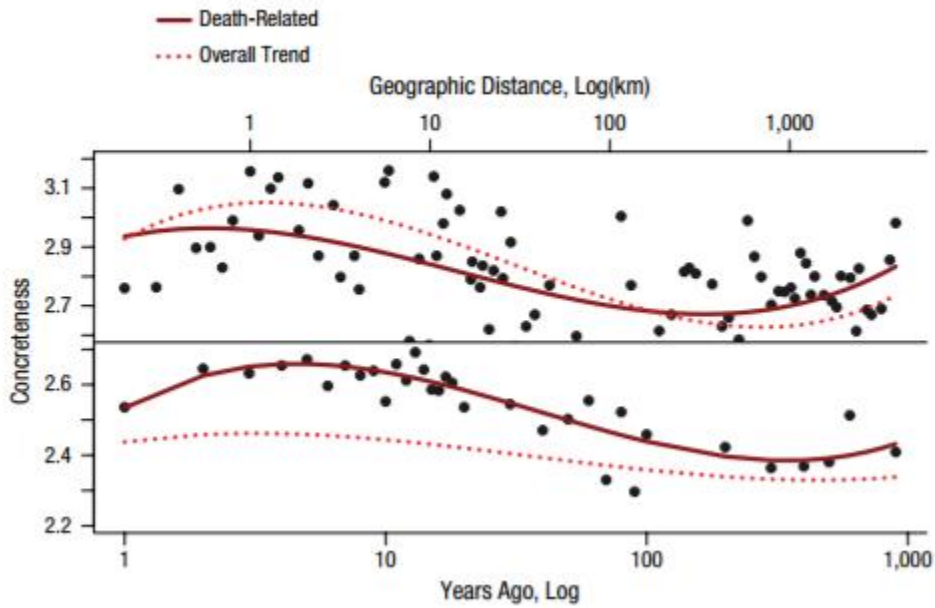


Fig. 4. Results from Study 4: scatterplots showing the mean concreteness of death-related Twitter messages regarding U.S. cities as a function of geographic distance (upper x-axis) and of death-related Usenet postings with the target phrase “years ago” as a function of log temporal distance (lower x-axis). Solid lines represent best-fitting regressions, and dotted lines replicate the concreteness of Twitter messages from Figure 1 and the concreteness of death-related Usenet postings with “years ago” from Figure 2a (upper and lower, respectively).