Cognition 152 (2016) 1–8

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Event construal and temporal distance in natural language

Sudeep Bhatia [a,*], Lukasz Walasek [b,*]

[a] Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Psychology and Behavioral Science Group at the Warwick Business School, University of Warwick, Coventry, United Kingdom

ARTICLE INFO

Article history:
Received 14 July 2015
Revised 28 February 2016
Accepted 11 March 2016

Keywords:
Construal level theory
Psychological distance
Natural language
Big data

ABSTRACT

Construal level theory proposes that events that are temporally proximate are represented more concretely than events that are temporally distant. We tested this prediction using two large natural language text corpora. In study 1 we examined posts on Twitter that referenced the future, and found that tweets mentioning temporally proximate dates used more concrete words than those mentioning distant dates. In study 2 we obtained all New York Times articles that referenced U.S. presidential elections between 1987 and 2007. We found that the concreteness of the words in these articles increased with the temporal proximity to their corresponding election. Additionally the reduction in concreteness after the election was much greater than the increase in concreteness leading up to the election, though both changes in concreteness were well described by an exponential function. We replicated this finding with New York Times articles referencing US public holidays. Overall, our results provide strong support for the predictions of construal level theory, and additionally illustrate how large natural language datasets can be used to inform psychological theory.

1. Introduction

With the ongoing digitization of information on the internet, it is now possible to access a large variety of natural language datasets. For a psychologist, these data offer an unprecedented gateway to study the formation of beliefs and attitudes, the dynamics of interpersonal relationships, and the preferences and behaviors of decision makers (Boyd & Crawford, 2012). Although still scarcely used in psychological science, these data have a potential to provide a unique and statistically powerful approach to evaluating models of human cognition, including models that are not specifically about language use.

In the following paper, we demonstrate how large corpora of natural language can be used to study the relationship between temporal distance and mental representations of events in the world. More specifically, we test the core assumption of the Construal Level Theory (CLT, Trope & Liberman, 2010) that the mental representation of a given object or event becomes less concrete and more abstract with increasing temporal distance. In two studies, we show that this hypothesis can be tested using online communication on Twitter and a large database of New York Times articles, by examining the concreteness of words used in these tweets and articles, and the distance between when they were written and when the events they pertain to occurred.

CLT has been used to describe the influence of psychological distance on the representations of physical objects (Liberman, Sagristano, & Trope, 2002), choice alternatives (Borovoi, Liberman, & Trope, 2010), events (Wakslak, Trope, Liberman, & Alony, 2006), consumer goods (Trope, Liberman, & Wakslak, 2007), actions (Liberman & Trope, 1998), and individuals (Rim, Uleman, & Trope, 2009). Psychological distance includes dimensions such as time, space, social distance and hypotheticality, although in the following paper we focus exclusively on the effect of temporal distance (Trope & Liberman, 2010). To illustrate, a person may describe a date as a "candlelight dinner" if the event occurred recently, but as a more abstract "romantic evening" if it occurred in the distant past. Higher level construals are not just impoverished versions of more proximal representations. Rather they involve different types of information, including information about the meaning of the object or action (Semin & Fiedler, 1988). According to CLT, the effect of psychological distance on our thought is functional as it allows us to plan for the future, learn from our mistakes or to communicate more efficiently.

The predictions of CLT find support in existing empirical studies (Soderberg, Callahan, Kochersberger, Amit, & Ledgerwood, 2015). The positive association between temporal distance and construal abstractness has been demonstrated in categorization tasks (Liberman et al., 2002), event descriptions (Liberman & Trope,

* Corresponding author.
E-mail addresses: bhatiasu@sas.upenn.edu (S. Bhatia), L.Walasek@warwick.ac.uk (L. Walasek).

http://dx.doi.org/10.1016/j.cognition.2016.03.011
0010-0277/

2008) and even at the level of visual perception (Förster, Friedman, & Liberman, 2004). Often simply imagining an event occurring in the distant future leads people to describe it using words with a high level of abstraction (Semin & Fiedler, 1988). Crucially, CLT has been shown to predict actual behavior, behavioral intentions, negotiation style, self-control, risk perception, and temporal discounting (see Trope et al., 2007 for a review). CLT thus attempts to provide a unifying framework for understanding the effect of distance on perceptual processes, social interactions, moral reasoning, consumer behavior, and even decision making under risk and uncertainty.

Most existing work on CLT uses laboratory experiments, in which psychological distance is induced or manipulated using hypothetical frames and primes (but see Magee, Milliken, & Lurie, 2010). For example, participants in these studies are often found to represent objects and events less concretely after being instructed to think of these items as distant (e.g., in spatial or temporal terms), once the information is presented as distant due to a specific reference to time or space (e.g., occurring in a distant future or happening far away), or when construal is induced by the question framing (e.g., thinking *why* rather than *how* an event occurred; Trope et al., 2007). While the exact methods of framing and priming vary greatly between the studies, many of these techniques require some form of artificial manipulation of psychological distance with reference to a given object or event. However, CLT also predicts that real-world psychological distance to an event should influence its representation in everyday thought and discourse. This is not easily tested in the lab.

In this paper we hope to test the predictions of CLT by analyzing the level of concreteness and abstraction of language in real-world settings. In study 1, we collected and analyzed millions of time-indexed posts on Twitter. Twitter is an excellent source of data – in 2015, it averaged 236 million active users (http://www.statista.com/statistics), who posted close to 500 million messages (tweets) per day. These tweets contain up to 140 characters and are shared among each user's social network (Reips & Garaizar, 2011). In this study we obtained a large number of tweets that referenced dates in the future, and were able to use these tweets to determine the concreteness of the language used to describe events at these dates. This allowed us to observe how psychological distance influences everyday discourse, and put the key assumptions of the CLT to a real-world test.

In study 2, we analyzed word concreteness in news articles using the New York Times (NYT) Annotated Corpus (Sandhaus, 2008). This corpus contains over 1.8 million NYT articles written between 1987 and 2007. Importantly for our purposes, these articles are tagged with keywords describing the topics of the articles. In this study we obtained all NYT articles written before and after the 1988, 1992, 1996, 2000, and 2004 US Presidential elections, which were tagged as pertaining to these elections. We subsequently tested how the concreteness of the words used in the articles varied as a function of temporal distance to the election they reference. We also performed this analysis with NYT articles referencing three popular public holidays. Unlike study 1 and prior work (such as Snefjella & Kuperman, 2015), study 2 allowed us to examine the influence of temporal distance in the past and in the future, while controlling for the exact time when specific events occurred. Recent findings show that future events are perceived as more proximal than equally distant past events (Caruso, Van Boven, Chin, & Ward, 2013). For example, events occurring 1 year in the future are rated as psychologically closer than events that occurred 1 year in the past. Our NYT dataset allows us to test whether this asymmetry is reflected in the abstractness with which various events are described.

2. Study 1: temporal distance and tweet concreteness

2.1. Methods

We first examined the effect of temporal distance on object representation by studying the relationship between the concreteness of the words used in tweets about objects and events at various points in the future. In particular, we obtained tweets that used the phrases "next week", "next month" or "next year", as well as tweets whose text mentioned the years "2015", "2016", "2017", and "2018". This was done using Twitter's data streaming feature, which allows researchers to download tweets as they are created. We filtered the Twitter data stream using the terms 'next week', 'next month', 'next year', '2015', '2016', '2017', and '2018' to obtain only the tweets that mentioned one of these phrases. We streamed Twitter over the course of one week in August 2014. The length of our data collection was set to one week as this provided enough time to obtain a very large number of tweets. Our collection was pruned to exclude retweets (that is, tweets that were copied and reposted).

We formalized word concreteness using a database of 40,000 English word ratings obtained by Brysbaert, Warriner, and Kuperman (2014). Brysbaert et al. collected these from over four thousand participants who were asked to rate different words on a 5-point scale based on how abstract or concrete the meanings of the words were to them. Using this database we scored each tweet on the average concreteness of its component words. The score for each tweet ranged from 1 for highly abstract to 5 for highly concrete. Tweets composed entirely of words absent from the Brysbaert et al. database were excluded from our dataset.

2.2. Results

Our final dataset for examining temporal distance consisted of 1,746,788 tweets that mentioned one of the three phrases or one of the four years of interest to us, and included at least one of the 40,000 words from the Brysbaert et al. database. The distribution of concreteness ratings for these tweets is displayed in the top panel of Fig. 1.

Construal level theory predicts that tweets referencing objects that are far away into the future should use relatively less concrete words than tweets referencing objects that are nearby in the future. Thus the average concreteness of tweets that mention "next week" should be higher than those that mention "next month", which in turn should be higher than those that mention "next year". Likewise the average concreteness of tweets that reference the years 2015–2018 should decrease with the year that they reference.

These predictions were supported by our data, which is summarized in Figs. 2 and 3, respectively. Particularly, tweets that mentioned "next week" had an average concreteness of 2.69 ($SD = 0.47$), tweets that mentioned "next month" had an average concreteness of 2.64 ($SD = 0.43$), and tweets that mentioned "next year" had an average concreteness of 2.48 ($SD = 0.39$). The differences between these three groups were statistically significant ($t = 36.65$, $p < 0.001$ for the difference between week and month, $t = 139.70$, $p < 0.001$ for the difference between month and year, and $t = 246.30$, $p < 0.001$ for the difference between week and year).

We obtained similar results for tweets that referenced the years 2015–2018. Particularly, tweets that referenced 2015 had an average concreteness of 2.79 ($SD = 0.58$), tweets that referenced 2016 had an average concreteness of 2.72 ($SD = 0.51$), tweets that referenced 2017 had an average concreteness of 2.70 ($SD = 0.46$), and tweets that referenced 2018 had an average concreteness of 2.61 ($SD = 0.40$). Overall, the concreteness of a tweet was lower if it referenced a year that was further into the future. Formally, we tested this using a linear regression of the effect of year on tweet