American Psychological Association

Journal of Personality and Social Psychology: Attitudes and Social Cognition

ISSN: 0022-3514

http://dx.doi.org/10.1037/pspa0000214

Subjective Likelihood and the Construal Level of Future Events: A Replication Study of Wakslak, Trope, Liberman, and Alony (2006)

Sofia Calderon, Erik Mac Giolla, Karl Ask, and Pär Anders Granhag
University of Gothenburg

C. J. Wakslak, Y. Trope, N. Liberman, and R. Alony (2006) examined the effect of manipulating the likelihood of future events on level of construal (i.e., mental abstraction). Over 7 experiments, they consistently found that subjectively unlikely (vs. likely) future events were more abstractly (vs. concretely) construed. This well-cited, but understudied finding has had a major influence on the construal level theory (CLT) literature: Likelihood is considered to be 1 of 4 psychological distances assumed to influence mental abstraction in similar ways (Trope & Liberman, 2010). Contrary to the original empirical findings, we present 2 close replication attempts ($N = 115$ and $N = 120$; the original studies had $N = 20$ and $N = 34$) that failed to find the effect of likelihood on construal level. Bayesian analyses provided diagnostic support for the absence of an effect. In light of the failed replications, we present a meta-analytic summary of the accumulated evidence on the effect. It suggests a strong trend of declining effect sizes as a function of larger samples. These results call into question the previous conclusion that likelihood has a reliable influence on construal level. We discuss the implications of these findings for CLT and advise against treating likelihood as a psychological distance until further tests have established the relationship.

Keywords: construal level theory, likelihood, hypotheticality, mental representation, replication

People can construe situations (e.g., going to a concert) in distinctly different ways. They can focus on the big picture (e.g., enjoying a cultural event) or the smaller pieces that make up the whole (e.g., talking loudly and ordering a drink). Construal level theory (CLT) provides an explanation for when and why people form more or less abstract mental representations of a situation (Trope & Liberman, 2010). It posits that increasing psychological distance promotes mental abstraction. Specifically, the further a situation is removed from the egocentric reference point of the here, now, self, and the real, the more abstract the mental representation will be. For example, a trip scheduled to take place in several months is expected to be represented at a high-level construal, producing thoughts related to the overarching purpose of the trip. As the trip approaches, however, it should be represented at a lower-level construal, producing more thoughts about concrete details, such as how to get to the airport. With this intuitive proposition, CLT has become one of the most influential social–cognitive theories during the last two decades. Social psychology handbooks dedicate entire chapters to CLT (Carlston, 2013; Fiske & Macrae, 2012; Kruglanski & Higgins, 2013) and seminal publications on the topic have been cited extensively. For example, as of May 2020, the literature review by Trope and Liberman (2010) had been cited 3,845 times on Google Scholar. This can be compared with the estimation that about 26 out of 100,000 publications in science overall will receive over 1,000 citations (Van Noorden, Maher, & Nuzzo, 2014).

Empirical studies support the assumption that psychological distance increases mental abstraction. In their work on temporal distance, Trope and Liberman (2003) found that events taking place in the far future were construed more abstractly than events in the near future. For example, participants who imagined having a yard sale in a year's time grouped objects relevant for the task in larger, more inclusive groups than participants imagining the action tomorrow. A similar link has been found between spatial distance and construal level: Henderson, Fujita, Trope, and Liberman (2006) found, for example, that participants divided a behavior imagined to take place in another city into broader categories than a behavior imagined to take place in their own city. Focusing on social distance, Liviatan, Trope, and Liberman (2008) found that participants described actions of persons similar to themselves in more "means-related" terms (indicating a more concrete action

Sofia Calderon, Erik Mac Giolla, Karl Ask, and Pär Anders Granhag, Department of Psychology, University of Gothenburg.

The current research was funded by a grant from the Swedish Research Council (VR; 2015-02144) awarded to Pär Anders Granhag. Thanks to the authors of the original study for their help with providing original material. Thanks to Hanni Beronius, Christel Lundin, and Linda Svedborg for their help with data collection.

Data and material for the two studies are found at the Open Science Framework https://osf.io/5x8dp/, where we have uploaded: Data sets for Studies 1 and 2, data set with statistics used for the new meta-analytic estimate, funnel plot, and *p*-curve analysis reported in the article (e.g., effect sizes, *p* values), stimulus material used in Studies 1 and 2, and complete survey with instructions and flow chart for the experimenter in Study 1 (note that these are in Swedish). Flow chart for the experimenter in Experiment 1.

Correspondence concerning this article should be addressed to Sofia Calderon, Department of Psychology, University of Gothenburg, P.O. Box 500, SE 405, 30 Haraldsgatan 1, 413 14 Göteborg, Sweden. E-mail: sofia.calderon@psy.gu.se

construal), whereas dissimilar persons' actions were described in more "ends-related" terms (indicating a more abstract action construal).

Of particular interest for the current study, however, is a study by Wakslak, Trope, Liberman, and Alony (2006). They hypothesized and found that subjective likelihood has a similar influence on mental abstraction as temporal, spatial, and social distance. They argued that an unlikely or hypothetical future event is, by its very definition, removed from direct experience. Hence, subjective likelihood should be associated with psychological distance. In seven experimental studies, Wakslak and colleagues manipulated subjective likelihood of future events and assessed participants' mental construal. In the majority of the experiments, the manipulation of likelihood consisted of telling participants that there was either a 5 or a 95% chance that they would perform a given action. Construal level was assessed using various measures common to this field of research, including the level of inclusiveness of objects and action segments (Experiments 1, 2, and 4), the generality of action descriptions (Experiment 3), performance on visual abstraction tests (Experiments 5 and 6), and performance on the Behavioral Identification Form (Experiment 7; see Burgoon, Henderson, & Markman, 2013, for an overview of measures of abstraction). Wakslak and colleagues consistently found statistically significant effects of likelihood manipulations on construal level in all experiments, with medium to large effect sizes (Hedges' g ranging from 0.39 to 1.34, with a weighted average of 0.63). Based on the consistent support for their predictions, they concluded that subjective likelihood was a strong candidate to be considered a form of psychological distance.

Meta-analytic findings in the CLT literature suggest that the effect of likelihood on construal level (Hedges' $g = 0.50$) is of similar size as the other three different forms of psychological distances (Hedges' gs of 0.49, 0.32, and 0.47) and has similar consequences for mental abstraction (Soderberg, Callahan, Kochersberger, Amit, & Ledgerwood, 2015).

The Importance of Replicating the Effect of Likelihood on Construal Level

It is important to replicate high-impact findings to produce reliable scientific knowledge (Asendorpf et al., 2013). The Wakslak et al. (2006) paper is a highly influential empirical paper, which has been extensively cited in scholarly work. In May 2020, it had 507 citations on Google Scholar, and there is a steady increase in the number of citations over time. These citations span over numerous research fields, including both basic psychology (e.g., Tomić, Tonković, & Ivanec, 2017) and applied psychology (e.g., Meyvis, Goldsmith, & Dhar, 2012). The original paper is published in a high impact journal (*Journal of Experimental Psychology: General*; Impact Factor = 4.10 in November 2018), which likely contributes to its high influence.

Despite the broad impact of Wakslak et al. (2006), likelihood remains an understudied dimension within CLT. A meta-analysis on the direct effect of psychological distance on construal level included 310 effects, only eight of which (0.03%) were effects of likelihood manipulations (Soderberg et al., 2015).[1] Seven of these eight effects were from the Wakslak et al. paper. Gilead, Liberman, and Maril (2012) conducted the remaining study, examining perceptual encoding of stimuli related to false or hypothetical events compared with a true event.[2] In a literature search during March of 2019, we did not find any additional relevant studies looking at the direct effect of subjective likelihood on construal level.

Despite the underrepresentation of the likelihood dimension, it is typically included in the definition of the term psychological distance (Liberman & Trope, 2014), and is, at least implicitly, given as much weight as the other dimensions that have received more research. In addition, the Wakslak et al. paper is the primary reference that authors use when presenting the psychological distance dimension of likelihood in review articles (e.g., Trope & Liberman, 2010), as well as in book chapters (e.g., Trope & Liberman, 2012). In other words, claims of likelihood as a psychological distance relies mainly on this one seminal paper.

Exacerbating the issue, the sample sizes of the seven experiments in the Wakslak et al. (2006) paper were consistently small. Sample sizes ranged from 20 to 95 participants ($M = 42.1$, $Mdn = 34$). Based on the mean sample size the studies had an overall power of 80% to detect a Cohen's d of 0.89. This is a large effect, to be compared with the overall effect size in social psychology of $d = 0.43$ (Richard, Bond, & Stokes-Zoota, 2003). Low-powered studies can be a serious issue, because effects that appear to be large can, in fact, be false positives or greatly overestimated (Nelson, Simmons, & Simonsohn, 2018).

The p-curve analysis can be used to examine signs of selective publication within a literature (Simonsohn, Nelson, & Simmons, 2014). A p-curve is a distribution of statistically significant p values, and its shape indicates whether statistically significant effects of $p < .05$ have evidential value or if they are more likely to be false positives. In a scenario where there is a true effect, the distribution should be right skewed; a larger proportion of p values should be expected around .01 than around .04. A left-skewed curve, on the other hand, is indicative of an effect with low replicability because such a distribution should only occur if significant p values are selectively reported or if there have been systematic attempts by researchers to bring p values below the .05 threshold through p-hacking (Simonsohn et al., 2014).

We used an online app (http://p-curve.com) to compute a p-curve based on the p values from previous studies on the effect of likelihood on construal level. The distribution of the p values was significantly left-skewed (see Figure 1). This suggests that there may have been selective reporting, publication bias, or p-hacking in favor of positive findings in the literature on the effect of likelihood on construal level, and questions the replicability of the studies that produced significant effects.

[1] CLT posits that psychological distance not only directly influences construal level, but that it also has so called downstream consequences on behavior.

[2] Under the assumption that likelihood is a form of psychological distance, Kahn and Björklund (2017) examined its effect on moral judgments—a so-called downstream consequence of mental abstraction. In line with the CLT-derived prediction that psychological distance promotes the moral content of a situation, Kahn and Björklund found a higher correspondence between moral values and judgments when participants judged a scenario described as being hypothetical compared with when it was described as real.

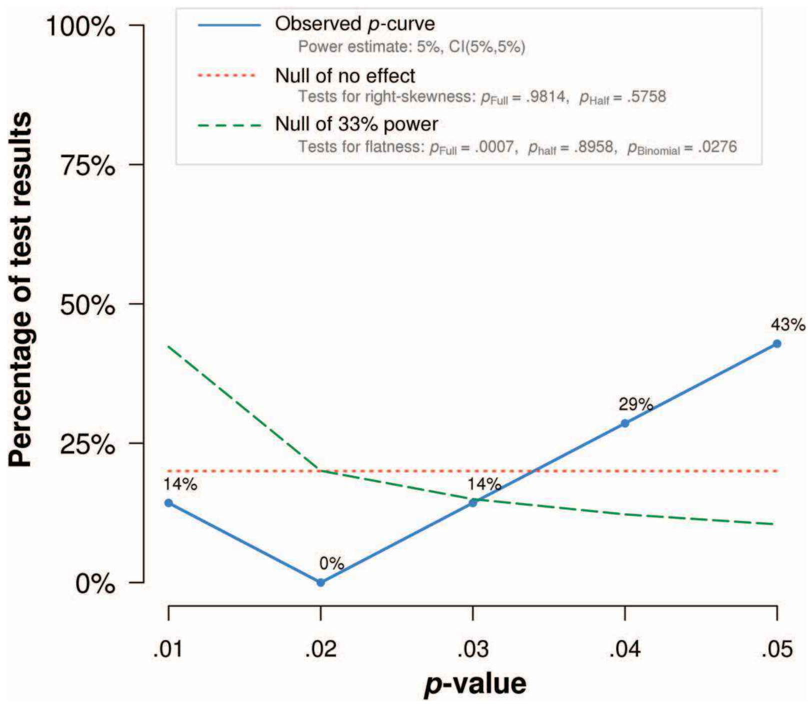

Figure 1. The *p*-curve analysis of previous experimental effects of likelihood on construal level. All seven effects were statistically significant ($p < .05$), two of which were $p < .025$. The *p*-curve indicates selective reporting, publication bias, or *p*-hacking in favor of positive findings. See the online article for the color version of this figure.

The Present Research

This study is a close replication of two out of the seven experiments in the Wakslak et al. (2006) study: Experiment 4 and Experiment 5. We manipulated participants' subjective likelihood of performing future tasks and then assessed how they mentally construed those tasks. More specifically, participants were told there was either a 5 or a 95% chance that they would later perform the tasks. Construal level was assessed using a categorization measure (i.e., a video segmentation task; Study 1), as well as two perceptual tests (i.e., visual abstraction; Study 2).

We preregistered our study decisions on the Open Science Framework (OSF) web page (Study 1: https://osf.io/x4va5; Study 2 https://osf.io/xumwa).[3] Specifically, we formally implemented the so-called Replication Recipe (Brandt et al., 2014) with an integrated template on OSF. The replication recipe suggests a number of criteria to be met for a study to be considered a close replication. For example, it includes specifying the extent to which the study properties are exact, close, or different from the original study when it comes to geographical location, instructions, stimuli, and procedure. We consider both our experiments to be close replications of the originals as they are as similar as possible on the specified criteria. The major differences were time (2006 vs. 2017/2018) and location (United States vs. Sweden) of data collection. The data and complete instructions used in both studies are available on the project page at the OSF website (https://osf.io/5x8dp/).

For Study 1, it was predicted that participants in the low likelihood condition would segment the video clip into fewer meaningful action units than those in the high likelihood condition. A Cohen's d greater than 0.52, which was the lower bound 95% confidence interval (CI) observed in the original study, was decided as the threshold for a successful replication of the original finding. A statistically significant effect was the criterion for a successful replication of the proposed effect of subjective likelihood on construal level. In Study 2, it was predicted that participants in the low likelihood condition would perform better at the two visual perception tests compared with the high likelihood conditions. A statistically significant crossover interaction effect was the successful replication criterion.

Study 1

Method

Participants. Participants were recruited for a study advertised as a study about behavior. To estimate an appropriate sample

[3] In the preregistration form for Study 2 we stated that the experimenter would be blind to condition. However, the fact that the experimenter manually timed participants' performance on the tasks—as was done in the original study—the experimenter was in fact not blind to the within-subjects condition (i.e., task). As for the second factor—the manipulation of likelihood—this was written on the same page as the timing instructions and we can, therefore, not rule out that the experimenter may have seen this instruction given to participants. However, there was nothing in the procedure that encouraged the experimenter to check anything more than the timing instructions.

size, we conducted an à priori power analysis based on the lower bound of the 95% CI (Cohen's $d = 0.52$) around the effect size observed in the original study, $d = 1.54$, 95% CI [0.52, 2.50] (Experiment 4; Wakslak et al., 2006).[4] There were 120 participants who were recruited, as this was needed to achieve 80% power for a two-sided between-groups t test when alpha level is set at 0.05. Five participants were excluded—one because of technical problems, one who had been in a previous similar study, three who did not follow the instructions—resulting in a final sample of 115 participants. A sensitivity power analysis revealed that the final sample size gave 80% power to detect an effect size of Cohen's $d = 0.53$ at an alpha level of 0.05. Eighty-one of the participants were female, 33 were male, and one did not specify their gender. Ages ranged from 19 to 78 years ($M = 32.1$, $Mdn = 26.0$, $SD = 14.1$). All participants received a lottery ticket (worth approximately $3) as compensation for their participation.

Procedure. After signing an informed consent form, participants were placed in front of a computer. They received one of two written instructions depending on their experimental condition, which had been randomly assigned before participants' arrival. The instructions were taken verbatim from the original study, but were translated to Swedish for our experiment. Participants in the low likelihood condition received the following instructions (with the high likelihood condition in parentheses):

> This experiment has two parts. You will now do the first part of the experiment, which involves watching a short video. In the video, you will see a woman performing the second part of the experiment. After watching the video I would like you to fill out a short questionnaire, and then you draw a note from the basket. In this basket there are 95 (5) notes saying "continue to the next part," and 5 (95) saying "thank you and goodbye." The note that you draw will determine whether you continue to the next part or not. As you can understand, there is a very high (low) probability that you will continue to perform the second part of the experiment.

The original authors were contacted but they no longer had access to the original video material. Therefore, we created a new video based on the description provided in the original article. In brief, the video consisted of a woman folding papers, drawing geometrical shapes on them, and counting the shapes. As in the original study, the video lasted approximately 5 min. One of the authors of the original paper verified that the video was highly similar to the original one (Rotem Alony, personal communication, October 17, 2017). Participants received the following instructions just before the segmentation task:

> What I would like you to do as you view this short movie is to segment the behavior into whatever actions seem natural and meaningful to you. Simply press the marked key [in original study: ENTER key] on the keyboard when, in your judgment, one meaningful action ends and another begins. These should be whatever actions seem natural and meaningful to you. There are no "right" or "wrong" ways to do this.

As in the original study, participants were reassured that the experiment was not about memorizing information, and that if they were to carry out the task they would be provided with the necessary information to complete it. Participants then viewed the video.

The number of indicated video segments was the dependent variable. A higher number of segments is indicative of a more concrete construal of the task, while a lower number of segments is indicative of a more abstract construal. After the segmentation task, participants filled out a brief questionnaire including demographic questions as well as the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). PANAS was not included in the original study. However, because positive affect has been shown to promote mental abstraction (Fredrickson & Branigan, 2005), we wanted to estimate mood as a potential confound in case of a successful replication.[5] Because this measure was collected after the dependent variable, it could not influence the results of primary interest.

Results and Discussion

The distribution of segments was positively skewed (skew = 2.49, $SE_{\text{skew}} = 0.23$) and, in accordance with the procedure of Wakslak et al. (2006), we transformed the variable using a logarithmic function (Log_{10}) before performing an independent t test. Failing to replicate the original effect, participants in the low likelihood condition ($M_{\text{transformed}} = 0.68$, $SD = 0.46$; $M_{\text{raw}} = 6.80$, $SD = 7.64$) did not segment the video clip into fewer units than did those in the high likelihood condition ($M_{\text{transformed}} = 0.59$, $SD = 0.48$, $M_{\text{raw}} = 6.12$, $SD = 8.99$), $t(113) = -1.05$, $p = .298$, $d = -0.19$, 95% CI [−0.56, 0.18]. In fact, the mean difference was in the opposite direction compared with the original experiment. We also conducted a Bayesian independent t test (one-sided) which revealed a Bayes Factor of 9.55 in favor of the null, which means that the data is over nine times more likely under the null hypothesis than under the alternative hypothesis. Based on these findings, we conclude that the finding in Experiment 4 of Wakslak et al. (2006) was not successfully replicated.

Study 2

Method

Participants. An à priori power analysis showed that only 28 participants were needed to detect an interaction effect of $\eta_p^2 = .071$—the lower bound of the 95% CI around the effect size for the 2×2 crossover interaction in the original study, $\eta_p^2 = .269$, 95% CI [.071, .442] (Wakslak et al., 2006, Experiment 5)—with 80% statistical power and an alpha of .05. Because resources were available to collect more data, we added extra participants. As in Study 1, 120 participants were recruited for the study. Seventy-three were women and 47 men, and ages ranged between 18 and 73 years ($M = 31.4$, $Mdn = 27.0$, $SD = 11.8$). Based on the targeted effect size, this sample size provided over 99% statistical

[4] We considered following the "small telescopes" approach to powering replication studies (Simonsohn, 2015), which proposes to use 2.5× the sample size of the original study. Because the original sample was very small (a total of 20 participants) this approach would have resulted in only 50 participants. Based on the available resources at our disposal, we decided instead to power the study for detecting the lower bound of the 95% confidence interval of the original effect.

[5] Participants' self-reported mood did not correlate with the dependent measure (i.e., number of behavior segments; $r = 0.03$, $p = 0.75$) and was, therefore, not included in any further analyses.

power with an alpha of .05 to detect the critical crossover interaction.

Procedure. Participants in the experiment were ostensibly recruited for a visual perception study. Upon arrival, and after signing an informed consent form, participants were given a booklet with written instructions. They were led to believe that they were going to perform one of two tests: the Snowy Pictures Test (SPT) or the Gestalt Completion Test (GCT; Ekstrom, 1976).[6] Both tests require the abstraction of visual information to identify an image (see Figure 2 for example pictures from both tests). These tasks were the same as the original study, and the authors of the original study were contacted to identify which items were used from each test (Cheryl Wakslak, personal communication, February 12, 2018).

All participants performed both tasks in what they believed were trial runs. Before performing the trial run, half of the participants were led to believe there was a 95% chance they would actually perform the SPT and a 5% chance they would actually perform the GCT (SPT likely/GCT likely condition). The remaining participants believed there was a 95% chance they would actually perform the GCT and a 5% chance they would actually perform the SPT (GCT likely/SPT unlikely). Participants were randomly assigned to conditions before arrival. We received the original instructions from the authors and translated them into Swedish. Before data collection, we asked a few Swedish-speaking psychology students to read them and evaluate their clarity. Based on their comments, slight alterations were made to the instructions to make them more comprehensible (see Appendix for the original and current study instructions). The manipulations were performed using the following instructions for the SPT likely/GCT unlikely condition (with the GCT likely/SPT unlikely condition in parentheses):

> Last semester we ran a study in which participants were presented with a series of incomplete pictures (pictures seemingly concealed with snow). The task was to name the objects depicted in the pictures (that were underneath the snow). We need a few more participants for that study. There is a small chance that you will be assigned to do that task instead. In other words, the majority of participants in this study will receive the snowy pictures task, while a few will receive the task involving incomplete pictures.

Before performing what they believed was the main task, all participants were given the opportunity to practice both the SPT and the GCT. Their performance on the tests during this practice session were in fact the dependent measures. The experimenter timed participants on these tasks, who had 3 min to finish the GCT and 90 s to finish the SPT. Performance on the SPT and GCT were calculated as the number of correctly identified objects on each test. The order of the tasks was counterbalanced. There were no indications of order effects.

A series of control questions were asked at the end of the experiment, taken verbatim from the original study. To check that the manipulation of subjective likelihood was successful, participants were asked two separate questions regarding how likely they were to receive the SPT (1 = *very unlikely*, 9 = *very likely*) and the GCT (1 = *very unlikely*, 9 = *very likely*). Participants were also asked how much they were looking forward to the experiment (1 = *very much*, 9 = v*ery little*) and how they were feeling at the moment (1 = *very bad*, 9 = *very good*).[7]

Results and Discussion

Participants in the SPT-likely condition perceived it as more likely to perform the SPT ($M = 7.52$, $SD = 2.03$, $Mdn = 8$) than those in the SPT-unlikely condition ($M = 2.98$, $SD = 2.30$, $Mdn = 2$), $t(118) = 11.46$, $p < .005$, $d = 2.09$, 95% CI [1.73, 2.46]. Also, participants in the GCT-likely condition rated it as more likely to receive the GCT ($M = 7.38$, $SD = 2.08$, $Mdn = 8$) than those in the GCT-unlikely condition ($M = 2.78$, $SD = 2.08$, $Mdn = 2$), $t(118) = 11.91$, $p < .005$, $d = 2.24$, 95% CI [1.88, 2.61].

In accordance with the analytic procedure of Wakslak et al. (2006), the number of correctly classified objects on the SPT and GCT tests were transformed into z-scores to allow for comparisons of performance across the two tests. We performed a 2×2 mixed analysis of variance (ANOVA), with test as the repeated measure and likelihood as the between-groups factor (SPT likely/GCT unlikely vs. SPT unlikely/GCT likely). Failing to replicate the original findings, the Test $\times$ Likelihood interaction was not statistically significant, $F(1, 118) = 0.04$, $p = .835$, $\eta_p^2 < .001$, 90% CI [.001, .015].[8] The significant crossover interaction from the original study and the nonsignificant interaction in the current study are illustrated in Figure 3. A Bayesian mixed ANOVA revealed a Bayes factor of 10.92 in favor of a null model over the alternative.

To explore simple effects, in accordance with the original study, an independent t test was performed on the raw scores of each test. There was no statistically significant difference in performance on

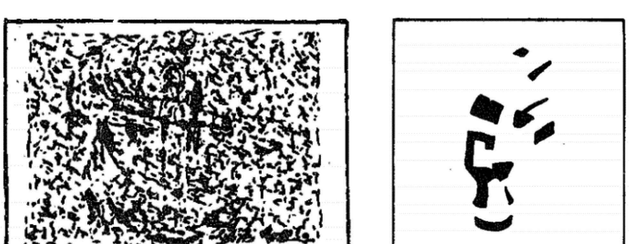

Figure 2. Example images from the Snowy Pictures Test (SPT; to the left) and the Gestalt Completion Test (GCT; to the right), depicting an anchor and a hammerhead, respectively.

[6] Copyright 1979 Educational Testing Service; www.ets.org. Figures from the Manual for Kit of Factor-Referenced Cognitive Tests are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this article is provided by the authors and no endorsement of any kind by Educational Testing Service should be inferred.

[7] The control questions did not correlate with the outcome measures: Participants' self-reported mood did not correlate with either SPT scores ($r = -.029$, $p = .756$) or the GCT scores ($r = .010$, $p = .916$); participants' reported level of looking forward to the experiment did not correlate with either SPT scores ($r = -.087$, $p = .343$) or GCT scores ($r = .054$, $p = .561$). These measures are not included in further analyses.

[8] The between-participants effect (i.e., SPT likely/GCT unlikely condition vs. GCT likely vs. SPT unlikely condition) on overall test performance was not statistically significant, $F(1, 118) = 0.26$, $p = .612$, $\eta_p^2 < .001$, 90% CI [.001, .036]. The within-participants effect of task (GCT vs. SPT) was nonexistent, $F(1, 118) = 0.00$, $p > .999$, $\eta_p^2 = .000$. The total lack of a within-participants effect was because of the z-transformation and the two task conditions, rendering a mean difference of 0.

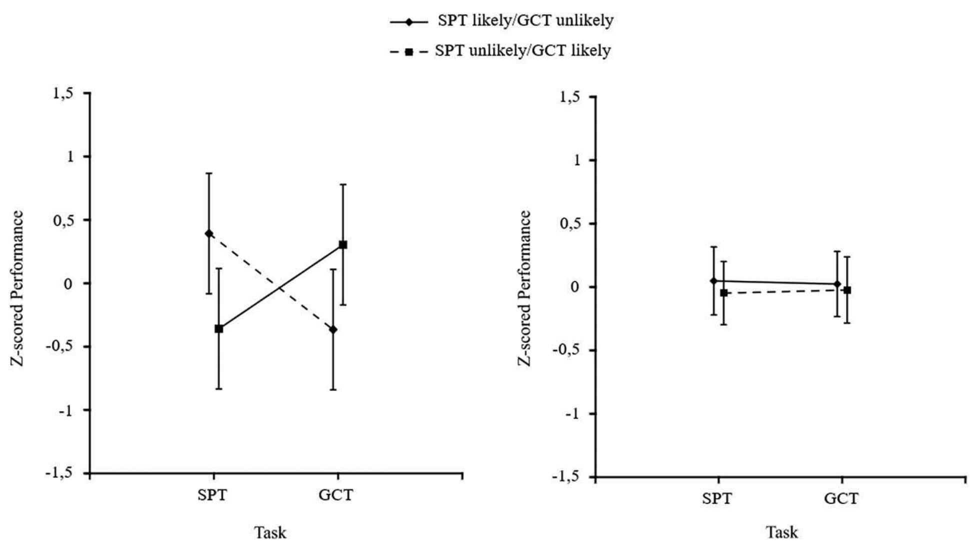

Figure 3. Statistically significant crossover interaction of original study (left) and nonsignificant interaction of the current replication study (right). Error bars represent 95% confidence intervals (CIs) of the means.

the SPT between participants in the low likelihood condition (M = 8.58, SD = 1.84) and those in the high likelihood condition (M = 8.77, SD = 1.99), $t(118) = -0.53$, $p = .600$, $d = -0.10$, 95% CI [−0.46, 0.26], $BF_{01} = 7.34$.[9] Also, no statistically significant difference was found in performance on the GCT between participants in the low likelihood condition (M = 3.12, SD = 2.10) and those in the high likelihood condition (M = 3.02, SD = 2.14), $t(118) = 0.26$, $p = .796$, $d = 0.05$, 95% CI [−0.31, 0.41], BF_{01} = 6.19.

Cumulative Evidence for the Effect of Subjective Likelihood on Construal Level

Because the original effects did not replicate in Study 1 or 2—while having considerably more participants than the original studies—we wanted to examine how sample size relates to the estimated size of the effect of likelihood manipulations on construal level in the literature at large. Using the package *ggplot2* (Wickham, 2016) in *R* (R Core Team, 2013), we created a funnel plot of effect sizes (Hedges' *g*) plotted against the standard error of the studies (see Figure 4). The standard error is commonly recommended as a proxy for study size in funnel plots (Sterne & Egger, 2001). The plot clearly shows that the effects become smaller as sample size increases. This indicates that previous experiments with small sample sizes may have overestimated the actual underlying effect. We conducted a new random effects meta-analysis using the package *metafor* (Viechtbauer & Viechtbauer, 2015) in *R*, where we included the current replication studies in addition to the effects from the Soderberg et al. (2015) meta-analysis (i.e., including the effects from Wakslak et al., 2006, and Gilead et al., 2012). It revealed an updated overall meta-analytic estimate of the effect of manipulated likelihood on construal level of Hedges' g = 0.44, 95% CI [0.18, 0.70], to be compared with the earlier estimate of Hedges' g = 0.50. A trim and fill analysis indicated that four studies are missing on the left side of the funnel plot. The adjusted trim-and-fill estimated effect size was Hedges' g = 0.25, 95% CI [−0.00, 0.52].

General Discussion

Wakslak et al. (2006) proposed that subjective likelihood may have an influence on construal level similar to that of temporal, spatial, and social distance. Specifically, they claimed that unlikely future events are more abstractly mentally represented than likely events. Over seven experiments—using multiple designs, measurements, and analyses—their data were consistent with this prediction. In contrast, the current substantially better powered replication attempts of two of the original experiments failed to find any such effect. Both Study 1 (measuring behavioral segmentation) and Study 2 (measuring visual abstraction) revealed effects close to and statistically indistinguishable from zero. It should be noted that the current replication failures were obtained in close replication attempts, meaning that the procedures, instructions, and stimuli material were as close to the original study as possible (Brandt et al., 2014).

Interpreting the Replication Outcomes

Failures to replicate a finding may occur for several reasons, only one of which is the possibility that the original finding was a false positive. Therefore, before accepting that our null findings reflect a genuine absence of the focal effect, it is important to assess the plausibility of the most salient alternative interpretations. These can be grouped into two major classes; the first relating to statistical power and precision, and the second relating to the problem of generalizing across variations in samples, settings, and procedures.

As for statistical considerations, our replication failures are not likely to be the result of limited power or precision. Our studies used samples almost six times bigger than those reported by Wakslak et al. (2006), yielding adequate power to

[9] The Bayes factors for the difference between the likelihood conditions on the SPT and GCT were based on Bayesian independent *t* tests (one-sided).

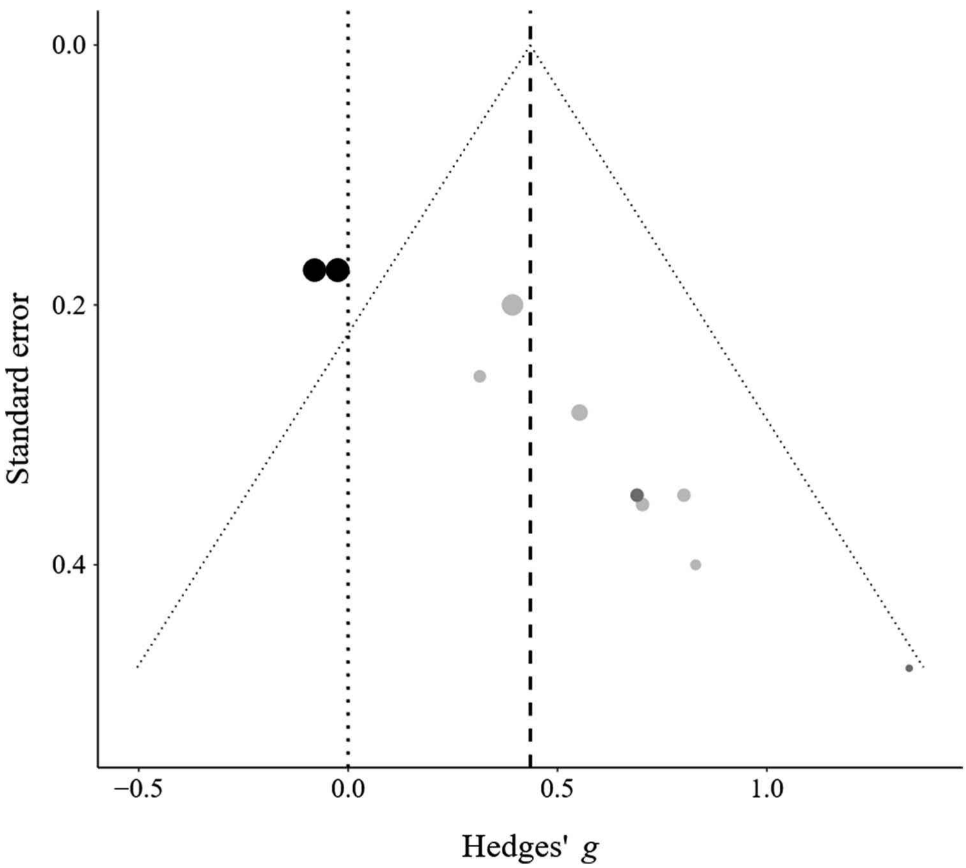

Figure 4. Funnel plot of the relationship between observed effect sizes in experiments examining the effect of likelihood on construal level (*x*-axis) and the standard error of the studies (*y*-axis). The effects are those included in the Soderberg et al. (2015) meta-analysis (Gilead et al., 2012; Wakslak et al., 2006), as well as the current replications. The dashed vertical line represents the new meta-analytic estimate. The dotted line is the point null effect. The black dots represent the current replication studies, the dark gray dots represent the original studies that were replicated, and the light gray dots represents the remaining studies on the topic. Sample size is indicated by the size of the dots (larger dots represent larger samples). Note that the effect size from the current Study 1 ($d = -0.08$; the smallest effect in the funnel plot) is based on the *raw* mean scores—as this was the procedure in the meta-analysis by Soderberg and colleagues—and not the Log-transformed scores that we report in the Results of Study 1 ($d = -0.19$).

detect effects substantially smaller than those reported in the original experiments. As a consequence, we were able to estimate effect sizes with considerably higher precision than was possible in the original research. Therefore, it seems more plausible that the original findings overestimated the underlying effect size, as opposed to the current study underestimating the effect size. Converging evidence for this conclusion comes from our analyses of the accumulated literature on the effect, which revealed two concerning patterns: First, the effect of likelihood on construal level becomes smaller as the samples used to study the effect get bigger. As far as we know, the current studies represent the two highest powered studies on the effect to date, corresponding to almost 45% of the combined samples from all previous studies (i.e., Gilead et al., 2012; Wakslak et al., 2006). Second, a *p*-curve analysis of the previously published findings revealed a pronounced overrepresentation of *p* values just below the conventional significance level of .05, suggesting the presence of selective reporting, publication bias, *p*-hacking, or any combination of the three (Simonsohn et al., 2014). In other words, it seems unlikely that the original findings accurately represent the true effect size.

In terms of between-study variation in samples and settings, the most obvious discrepancies between the original and the current experiments are the place (United States and Israel vs. Sweden) and the time (2006 vs. 2018) of data collection. The critical question is whether these differences could plausibly account for the distinctly different observed results. On the one hand, previous research has shown that some cultural variation exists in baseline levels of psychological distance. For example, Russians have been shown to self-distance (i.e., taking the perspective of others) to a higher degree than Americans (Grossmann & Kross, 2010), suggesting cultural differences in social distance. Relatedly, Messervey (2008) found a weaker association between temporal distance and construal level among Chinese than among Canadian participants. On the other hand, the United States, Israel, and Sweden are all examples of WEIRD (Western, educated, industrialized, rich, and democratic; Henrich, Heine, & Norenzayan, 2010) countries, and we are hard pressed to come up with salient cultural differences that could explain the presence of the original effects in both American and Israeli samples, but not in Swedish samples. Moreover, the recent Many Labs 2 study (Klein et al., 2018), attempting to replicate 28 classic and contemporary psychology findings in 36 countries and territories, showed that variations in sample and setting (including WEIRDness) provide little explanatory power for failures to replicate. Further speaking against the possibility that cultural differences are responsible for the current replication failure, both the English and the Swedish languages contain expressions that reveal a semantic relationship between the concepts of likelihood and distance (e.g., "near certainty" and "remote possibility").

Although the procedural differences between the original studies and our replication attempts were kept to a minimum, slight variations were inevitable. Of particular note, the original authors no longer had access to the video material used for the segmentation task (Experiment 1) and, as a consequence, we had to create a new video. Although one of the original authors deemed it highly similar to the original, the number of segments identified by participants in the current Experiment 1 ($M_{\text{low likelihood}} = 6.8$ vs. $M_{\text{high likelihood}} = 6.1$) differed notably from the number of segments identified in the original study ($M_{\text{low likelihood}} = 14.8$ vs. $M_{\text{high likelihood}} = 49.4$). One possible explanation for this disparity is that the current material contained a smaller number of behavioral units than the original material, which limited the number of observable segments. Although we deem this unlikely, given that we modeled the material on the description of the original material and had the similarity confirmed by one of the original authors, this possibility cannot be excluded entirely. A second possibility is that undocumented instructions provided to participants in the original study encouraged more frequent segmentation responses. As we have access only to Wakslak et al.'s documented instructions (and used them to instruct our participants) we cannot assess the plausibility of this explanation. However, it would seem that quite substantial additions to the documented instructions would be necessary to produce very large differences in participants' segmentation behavior. A third possibility is that our participants' characteristics (e.g., being Swedish and more heterogeneous than in the original study) contributed to an overall higher level of construal. This seems unlikely, however, given that a study with participants recruited from the same participant pool (Calderon, Ask, Mac Giolla, & Granhag, 2019) reported levels of mental construal—estimated using the Behavioral Identification Form (Vallacher & Wegner, 1987)—very similar to those reported by a number of American university undergraduate samples (Vallacher & Wegner, 1989), as well as those reported in CLT studies from the same laboratory as the Wakslak et al. (2006) study (Fujita, Henderson, Eng, Trope, & Liberman, 2006, Experiment 1; Liviatan et al., 2008, Experiment 1). It should also be noted that there is substantial spread in the number of segments found in the previous literature using the video segmentation task. For example, Newtson (1973, Experiment 2) found an average of 3.3 segments/minute (i.e., 16.7 segments in a 5-min video), while Henderson et al. (2006) found 9.6 segments/minute (i.e., 12.3 segments in a 76 s video) and Wilder (1978) found 16.2 segments/minute (i.e., 80.9 segments in a 5-min video). Given the large heterogeneity in segmentation behavior across studies, our observed level (1.3 segments/minute) is not remarkable.

While none of the above alternative explanations can be conclusively refuted, they cannot plausibly account for the fact that the massive effects ($d = 1.54$ and $\eta_p^2 = .27$) observed by Waklsak et al. in very small samples failed to replicate, or even assume the same direction, in samples almost six times the size ($d = -0.19$ and $\eta_p^2 < .001$). We believe that the most parsimonious explanation for the discrepant findings is that the current high-powered replication attempts provide more precise and more accurate effect size estimates.

Theoretical Implications

The CLT model suggests that psychological distance has a direct effect on construal level which, in turn, influences behavior (so-called "downstream consequences"; Soderberg et al., 2015). In other words, the direct effect of psychological distance on mental abstraction is posited as the psychological mechanism behind behavioral predictions derived from the theory. Previous studies have found so-called indirect ("downstream") effects of likelihood or hypotheticality on behavior (Todorov, Goren, & Trope, 2007; Kahn & Björklund). Insofar as the current study sheds doubt on the likelihood-construal level link, these previous findings may be because of some mechanism other than mental abstraction. It is also worth noting that previous studies have shown that people associate likelihood with the other proposed psychological distance dimensions (Bar-Anan, Liberman, & Trope, 2006; Bar-Anan, Liberman, Trope, & Algom, 2007; Fiedler, Jung, Wänke, & Alexopoulos, 2012; Maglio, Trope, & Liberman, 2013). Given the current findings, it is possible that such associations exist without likelihood necessarily having effects on construal level similar to those of the other distance dimensions.

There is reason to assume that likelihood as a form of psychological distance is particularly elusive to study. People are generally poor at dealing with probabilities and make erroneous judgments of likelihood in a variety of situations (Tversky & Kahneman, 2002). Therefore, it could be difficult to successfully manipulate subjective likelihood, making experimental manipulations such as those typically employed in CLT studies (e.g., 5 vs. 95% chance of performing a task) ineffective in changing people's perception of psychological distance. Trope and Liberman (2010) reasoned in similar terms as they proposed that likelihood might be the least prominent distance dimension of the four: the cognitive skill of thinking in terms of probabilities is acquired at an older age compared with the other dimensions in the CLT framework, such as spatial distance, which is learnt at a very young age. The Bar-Anan et al. (2007) study provides some indication that hypotheticality and distance are not necessarily closely related. In three experiments, Bar-Anan et al. examined the automatic activation of distance-related concepts by exposure to words related to hypotheticality. A close examination of the results shows that two analyses produced statistically significant associations, whereas an attempt to replicate one of these effects did not succeed. All other eight experiments on the remaining three dimensions produced statistically significant effects in the predicted direction. Hence, that paper provides ambiguous evidence for the link between distance and hypotheticality. Future studies should further examine the assumed association between the concepts as this is crucial for the viability of likelihood as a psychological distance in the CLT model.

It should also be noted that the CLT literature usually treats manipulations of likelihood (high vs. low) and hypotheticality (real vs. imagined) as interchangeable manipulations of the same underlying construct. However, these manipulations may not be equivalent, and it could be argued that our null-findings only call into question previous findings resulting from likelihood manipulations. Even if so, the empirical evidence for a direct effect of manipulating the hypotheticality dimension is nonetheless weak. Only Gilead et al. (2012) have examined and found support for a direct effect of a hypotheticality manipulation on construal level, whereas Kahn and Björklund (2017) studied its effects on downstream consequences. Thus, based on the current state of the evidence, there is currently very little evidence that either likelihood manipulations or hypotheticality manipulations influence construal level.

A caveat should be mentioned regarding the outcome measures used for assessing construal level in the current research. The two studies used three dependent measures of construal level (i.e., one behavior segmentation task and two visual perception tests). This by no means exhausts the range of dependent variables one could use to measure mental abstraction (Soderberg et al., 2015). Hence, it would be worthwhile to further explore the effect of likelihood manipulations using alternative outcome measures. Nevertheless, these measures have been noted in the literature as common and successful measures of mental abstraction (Burgoon et al., 2013; Trope & Liberman, 2010). Moreover, the exact same measures were used in the original study by Wakslak et al. (2006) and we have no reason to suspect that they would have been any less sensitive in the current replication attempts.

Conclusions

Our failed replications, in combination with the limited statistical power of the original studies, the outcome of the *p*-curve analysis, and the inverse relationship between sample size and observed effect size, greatly undermine the strength of the evidence in support of likelihood as a fourth psychological distance dimension within construal level theory. We attempted to replicate two of the largest direct effects previously found for likelihood manipulations on construal level. If these original effect estimates were at least somewhat accurate, they would almost certainly have been detected in our more high-powered tests. This indicates that the original findings may have been false positives or, at the very least, gross overestimations of the true effect. Therefore, we recommend caution when interpreting past research demonstrating the influence of subjective likelihood on mental abstraction, and advise against treating likelihood as a psychological distance until further tests have established the relationship.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27,* 108–119. http://dx.doi.org/10.1002/per.1919

Bar-Anan, Y., Liberman, N., & Trope, Y. (2006). The association between psychological distance and construal level: Evidence from an implicit association test. *Journal of Experimental Psychology: General, 135,* 609–622. http://dx.doi.org/10.1037/0096-3445.135.4.609

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. http://dx.doi.org/10.1037/0096-3445.136.4.610

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50,* 217–224. http://dx.doi.org/10.1016/j.jesp.2013.10.005

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. http://dx.doi.org/10.1177/1745691613497964

Calderon, S., Ask, K., Mac Giolla, E., & Granhag, P. A. (2019). The mental representation of true and false intentions: A comparison of schema-consistent and schema-inconsistent tasks. *Cognitive Research: Principles and Implications, 4,* 29. http://dx.doi.org/10.1186/s41235-019-0173-4

Carlston, D. E. (2013). *The Oxford handbook of social cognition*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199730018.001.0001

Ekstrom, R. B. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Fiedler, K., Jung, J., Wänke, M., & Alexopoulos, T. (2012). On the relations between distinct aspects of psychological distance: An ecological basis of construal-level theory. *Journal of Experimental Social Psychology, 48,* 1014–1021. http://dx.doi.org/10.1016/j.jesp.2012.03.013

Fiske, S. T., & Macrae, C. N. (2012). *The SAGE handbook of social cognition*. Los Angeles, CA: Sage.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19,* 313–332. http://dx.doi.org/10.1080/02699930441000238

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. http://dx.doi.org/10.1111/j.1467-9280.2006.01698.x

Gilead, M., Liberman, N., & Maril, A. (2012). Construing counterfactual worlds: The role of abstraction. *European Journal of Social Psychology, 42,* 391–397. http://dx.doi.org/10.1002/ejsp.1862

Grossmann, I., & Kross, E. (2010). The impact of culture on adaptive versus maladaptive self-reflection. *Psychological Science, 21,* 1150–1157. http://dx.doi.org/10.1177/0956797610376655

Henderson, M. D., Fujita, K., Trope, Y., & Liberman, N. (2006). Transcending the "here": The effect of spatial distance on social judgment. *Journal of Personality and Social Psychology, 91,* 845–856. http://dx.doi.org/10.1037/0022-3514.91.5.845

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83. http://dx.doi.org/10.1017/S0140525X0999152X

Kahn, D. T., & Björklund, F. (2017). Judging those closest from afar: The effect of psychological distance and abstraction on value–judgment correspondence in responses to ingroup moral transgressions. *Peace and Conflict, 23,* 153–161. http://dx.doi.org/10.1037/pac0000248

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., . . . Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1,* 443–490. http://dx.doi.org/10.1177/2515245918810225

Kruglanski, A. W., & Higgins, E. T. (2013). *Social psychology: Handbook of basic principles*. New York, NY: Guilford Press Publications.

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences, 18,* 364–369. http://dx.doi.org/10.1016/j.tics.2014.03.001

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. http://dx.doi.org/10.1016/j.jesp.2008.04.007

Maglio, S. J., Trope, Y., & Liberman, N. (2013). Distance from a distance: Psychological distance reduces sensitivity to any further psychological distance. *Journal of Experimental Psychology: General, 142,* 644–657. http://dx.doi.org/10.1037/a0030258

Meyvis, T., Goldsmith, K., & Dhar, R. (2012). The importance of the context in brand extension: How pictures and comparisons shift consumers' focus from fit to quality. *Journal of Marketing Research, 49,* 206–217. http://dx.doi.org/10.1509/jmr.08.0060

Messervey, D. L. (2008). Time and time again: Cultural differences in construal levels. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69,* 2007.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69,* 511–534. http://dx.doi.org/10.1146/annurev-psych-122216-011836

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28,* 28.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7,* 331–363. http://dx.doi.org/10.1037/1089-2680.7.4.331

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26,* 559–569. http://dx.doi.org/10.1177/0956797614567341

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143,* 534–547. http://dx.doi.org/10.1037/a0033242

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. http://dx.doi.org/10.1037/bul0000005

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54,* 1046–1055. http://dx.doi.org/10.1016/S0895-4356(01)00377-8

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Todorov, A., Goren, A., & Trope, Y. (2007). Probability as a psychological distance: Construal and preferences. *Journal of Experimental Social Psychology, 43,* 473–482. http://dx.doi.org/10.1016/j.jesp.2006.04.002

Tomić, I., Tonković, M., & Ivanec, D. (2017). Effects of psychological distance and need for cognitive closure on impression formation. *Journal of General Psychology, 144,* 1–15. http://dx.doi.org/10.1080/00221309.2016.1258385

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110,* 403–421. http://dx.doi.org/10.1037/0033-295X.110.3.403

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117,* 440–463. http://dx.doi.org/10.1037/a0018963

Trope, Y., & Liberman, N. (2012). Construal level theory. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 118–134). London, UK: SAGE Publications. http://dx.doi.org/10.4135/9781446249215.n7

Tversky, A., & Kahneman, D. (2002). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 19–48). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511808098.003

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological review, 94,* 3–15.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social psychology, 57,* 660–671.

Van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature, 514,* 550–553. http://dx.doi.org/10.1038/514550a

Viechtbauer, W., & Viechtbauer, M. W. (2015). Package 'metafor'. The comprehensive R archive network. Retrieved from http://cran.r-project.org/web/packages/metafor/metafor.pdf

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. http://dx.doi.org/10.1037/0096-3445.135.4.641

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.

Wilder, D. A. (1978). Effect of predictability on units of perception and attribution. *Personality and Social Psychology Bulletin, 4,* 281–284. http://dx.doi.org/10.1177/014616727800400222

(*Appendix follows*)

Appendix

Original Study Instructions and Instructions Used in the Current Study 2

Original instructions	Replication instructions
This experiment is the third experiment we are running as part of a line of work focusing on how people perceive pictures. The experimental task you will be given consists of a series of pictures that are partially concealed by snow. Your task will be to recognize the objects that are underneath the snow.	This experiment is the third experiment we are running as part of a line of work focusing on how people perceive pictures. You will be presented with a series of pictures that are partially concealed by snow. Your task will be to recognize the objects that are underneath the snow.
Before beginning, we wish to mention that a small percentage of people signed up for this experiment will be randomly assigned to complete a different task. During the summer semester we ran a study in which participants were presented with incomplete pictures and were asked to name the objects the pictures represented. We decided that we needed a few more participants for that study, so a small number of random people will be assigned to be in that experiment instead of this one. However, the vast majority of you will participate in the snowy picture task described.	We also want to inform about the following: Last semester we ran a study in which participants were presented with incomplete pictures. The task was to identify the objects that the pictures represented. We need a few more participants for that study. There is a small chance that you will be assigned to do that task instead. In other words, the majority of participants in this study will receive the snowy pictures task, while a few will receive the task with incomplete pictures.
The actual experiment will consist of 100 trials on the computer. Before you start the actual experimental session, you will have a chance to practice a paper and pencil version of both tasks. This is just a chance to familiarize yourself with the way these tasks work, and your performance on the paper and pencil practice trials won't be examined. After finishing the practice session, you will move to the computer where you will participate in the actual experimental session.	The actual experiment consist of 100 pictures from one of the tasks described above. The experiment will be performed on a computer. Before you start the actual experimental session, you will have a chance to practice a paper and pencil version of both tasks. This is a chance to familiarize yourself with the way these tasks work. Your performance on the paper and pencil practice trials won't be examined. After finishing the practice session, you will move to the computer where you will be randomly assigned one of the two tasks that you practiced.
Thanks in advance for your participation.	Thanks in advance for your participation.

Received April 30, 2019
Revision received February 3, 2020
Accepted April 22, 2020 ■