Effects of Psychological Distance on Mental Abstraction: A Registered Report of Four Tests of Construal Level Theory

Proposing/Lead Authors: Sofia Calderon, Erik Mac Giolla, Karl Ask, Timothy J. Luke
Contributing Authors: See published version or the author information file available here: https://osf.io/63wyr

Multi-lab direct replication of

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, *75*(1), 5-18. Study 1.

Fujita, K., Henderson, M., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science*, *14*, 278-282. Study 1.

Plus two paradigmatic replications of the effect of social distance and hypothetical distance on construal level.

Citation: Calderon, S., Mac Giolla, E., Ask, K., … Luke, T. J. (2025). Effects of psychological distance on mental abstraction: A registered report of four tests of construal level theory. *Advances in Methods and Practices in Psychological Science*.

Supplemental Material: https://osf.io/ra3dp/, https://github.com/RabbitSnore/CLIMR. The repositories contain the data, the full study protocols and changes made from the original studies, Qualtrics Survey Files (.qsf) to recreate the surveys, procedure videos, and R code for reproducing the analyses.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Acknowledgements: This project was supported by a grant from the Swedish Research Council awarded to the second author (grant nr. 2019-03037). We would like to thank all original authors who have provided materials for this project.

Address correspondence to: Sofia Calderon, sofia.calderon@psy.gu.se

Abstract

Construal level theory (CLT) proposes that psychological distance influences the level of abstraction at which something is mentally construed: things perceived as less probable (likelihood) or further away from the here (spatial distance), now (temporal distance), or self (social distance) are thought about more abstractly. This international multi-lab study tests four basic hypotheses derived from core assumptions of CLT and explores potential moderators and boundary conditions of the effects. Participants (N = 11,775) from 27 countries and regions were randomly assigned to one of four experimental protocols focused on different types of psychological distance (temporal, spatial, social, or likelihood), and each experiment manipulated psychological distance (close vs. distant). The protocols for temporal distance (N = 2,941) and spatial distance (N = 2,973) were direct replications of Liberman and Trope (1998, Study 1) and Fujita et al. (2006, Study 1), respectively. The remaining two protocols were paradigmatic replications, applying to social distance (N = 2,926) and likelihood (N = 2,936). The effects of psychological distance on construal level for the four present studies were (original effects within parentheses; positive effects are consistent with hypotheses): $d_{temporal}$ = 0.08, 95% CI [0.003, 0.16] (cf. d = 0.92); $d_{spatial}$ = 0.04, 95% CI [-0.03, 0.11] (cf. d = 0.55); d_{social} = -0.27, 95% CI [-0.34, -0.19]; and $d_{likelihood}$ = 0.03, 95% CI [-0.05, 0.11]. Pretests indicated that valence and abstraction were confounded in response options on the outcome measure. Controlling for this confound eliminated the hypothesis-inconsistent effect of social distance, d = 0.006, 95% CI [-0.05, 0.07]. These findings provide limited evidence for the predictions of the theory and present a critical challenge for CLT.

Keywords: construal level theory, mental abstraction, psychological distance, replication, multi-lab

Effects of Psychological Distance on Mental Abstraction: A Registered Report of Four Tests of Construal Level Theory

The mind's ability to represent the world in concrete terms or as abstract concepts is a fundamental aspect of human cognition. This ability is central to understanding processes underlying, for instance, prejudice, judgment and decision making, and problem solving (Burgoon et al., 2013). Being able to predict how objects, people, and events are mentally represented is therefore essential to understanding how people interact with the world around them. Construal level theory (CLT) is a framework developed to explain when and why the mind construes objects and events in more concrete (low-level) or abstract (high-level) terms (Trope & Liberman, 2010).

According to CLT, how an object is construed depends on how psychologically distant it is perceived to be. CLT suggests that as perceived psychological distance increases, objects and events should be represented at a higher construal level. That is, they should be represented in more abstract, simple, and decontextualized terms, compared to objects and events perceived as psychologically close (Trope & Liberman, 2003). CLT proposes four types of psychological distance: temporal, spatial, social, and likelihood (or hypothetical) distance. Specifically, objects and events can be perceived as close or distant in time or space, people can be perceived as close (e.g., similar) or distant (e.g., dissimilar) to the self, and events can be perceived as close (likely/actual) or distant (unlikely/hypothetical) from the real world. Increased distance on any of these dimensions should lead to higher level, more abstract, construals (for a review see Trope & Liberman, 2010). We will refer to this as the *direct* effect of psychological distance on construal level (as opposed to indirect, downstream effects on other variables).

As an example, let us consider how temporal distance may influence construal level. Imagine that you are to attend a friend's wedding in a year's time. According to CLT, at the current point in time, you are likely to think about the event in abstract and decontextualized

terms. Your representation of the wedding should be schematic, focusing on typical and core aspects, such as well-dressed people and the value of celebrating a couple's love for each other. However, as the day approaches, your representation of the wedding should become more contextualized and specific. The day before the wedding your thoughts may be more on details like what to wear and how to get from your hotel to the wedding ceremony.

The above example describes a few ways in which a difference in construal level may manifest itself; specifically, that greater distance increases people's tendency to think of actions in terms of their purpose instead of their concrete steps of implementation. The CLT literature proposes several other manifestations of higher-level construal—for example, an increased breadth with which people categorize objects and an increased tendency to focus on the whole rather than the parts. A large number of dependent measures have been developed to assess construal level and mental abstraction (for a list of commonly used measures of abstraction, see Burgoon et al., 2013).

Apart from the direct effect of psychological distance on construal level, CLT also proposes downstream consequences. These are secondary effects of psychological distance on behavior. More specifically, the theory proposes an indirect path, such that level of mental construal mediates the effect of psychological distance on behavior. For example, previous studies have found that psychological distance influences performance predictions (e.g., of one's ability to perform a task in the near or distant future; Nussbaum, et al., 2006), evaluations (e.g., of an essay written by someone similar or dissimilar to oneself; Liviatan et al., 2008), and behavioral intentions (e.g., the number of hours one is willing to volunteer in the near or distant future; Eyal et al., 2009). These findings have been interpreted as the result of varying levels of mental construal. Although research on downstream consequences constitutes a large part of the CLT literature, the topic lies outside the scope of the current research. Here, we focus specifically on direct effects of psychological distance on construal level.

By now, hundreds of experiments that test predictions from CLT have been published (for a bibliometric overview, see Adler & Sarstedt, 2021). In the most comprehensive meta-analysis to date, Soderberg et al. (2015) concluded that the existing literature provided support for a medium-sized effect of psychological distance on both mental abstraction and downstream consequences. In addition to this large body of work, converging evidence for the theory has been reported in other areas of enquiry. This includes effects of social distance on person-perception, temporal effects on memory, and the relationship between power differences and social distance (Trope & Liberman, 2010). Consider, for example, the correspondence bias in person-perception research, which is the increased tendency to ignore situational information and draw inferences about an actor's stable traits when judging others' (vs. one's own) behavior (Gilbert & Malone, 1995). From a CLT perspective, this is because others are construed more abstractly than the self, resulting in a greater focus on abstract, decontextualized dispositions (Trope & Liberman, 2010).

Despite the vast body of work on CLT, replication studies are rare. Independent replications are crucial in order to obtain accurate estimates of effect sizes, to uncover potential moderators, and to determine whether the effects are replicable outside the original labs (Nosek & Errington, 2020; Simons, 2014). The few extant replication attempts provide a mixed picture of the replicability of CLT findings. These studies have produced either (a) mixed results that both replicated and contradicted the original findings (Luke et al., 2021; Žeželj & Jokić, 2014); (b) non-significant results or results in the opposite direction of the original findings (Calderon et al., 2020; Gong & Medin, 2012; McCarthy et al., 2018); or (c) estimates of effect sizes in the expected direction, but substantially smaller than the originally reported effects (Sánchez et al., 2021). Furthermore, with the exception of Calderon et al. (2020) and Sánchez et al. (2021), these replication studies have focused on downstream consequences of construal level; that is, the effect of psychological distance on behavior. Studies on direct effects are central to CLT's

assumption that construal level is the mechanism driving the influence of psychological distance on behavior.

In a recent unpublished preprint, Maier et al. (2024) reanalyzed the meta-analytic data of Soderberg et al. (2015) using novel robust Bayesian meta-analytic techniques (Bartos et al., 2021). This reanalysis estimated a bias-corrected effect size near zero for direct effects of psychological distance on construal level. Additionally, Maier et al. found that the rate of positive results in the CLT literature far exceeds that which would be expected given the average statistical power of studies in the literature. These signs of bias, in conjunction with the limited number of independent replication attempts, highlight the need for powerful tests of the robustness and boundary conditions of CLT's hypotheses.

The Present Research

The present research adds to the CLT literature by conducting direct and paradigmatic replications of the direct effects of four psychological distances on construal level. Using an international multi-lab approach, we directly replicated the following two studies: Liberman and Trope (1998, Study 1; *temporal distance*) and Fujita et al. (2006, Study 1; *spatial distance*). In addition, the experimental paradigms used in the above studies were extended to test the two remaining distances—*social distance* and *likelihood*. All four studies used the Behavior Identification Form (BIF; Vallacher & Wegner, 1989)—which is the most widely used measure of abstraction in the literature (Burgoon et al., 2013)—as the main dependent variable.

Our primary analyses examined whether the effects in the four studies were in the direction predicted by CLT. In addition, for the two direct replications, we examined whether the replication effects were consistent with the original results in terms of direction and size of the effect. Furthermore, the large scale of the project allowed us to examine potential moderators, thereby addressing a critical research gap noted by others in the field (Soderberg et al., 2015). Among other things, we examined whether the effects are contingent on the mode

of data collection (online vs. in-lab) and regional variations. The moderator analyses were aimed at identifying potential boundary conditions of the tested effects, which may prompt further specification of the theory and/or revision of its hypotheses.

Method

Identification of Suitable Studies

To be a candidate for replication for the current project, a study should have (a) experimentally manipulated one, and only one, form of psychological distance and (b) used a direct measure of construal level as the dependent variable. To identify suitable studies, we screened experiments included in the Soderberg et al.'s (2015) meta-analysis on direct effects of psychological distance on construal level. Out of an original 134 experiments, 47 were excluded because, in our view, they did not examine a direct effect of psychological distance on construal level (e.g., we did not consider ratings of the feasibility vs. desirability of an outcome as a direct effect, in line with Liberman et al., 2002). The remaining 87 experiments were screened further. For details on the screening procedure see https://osf.io/tpf6v/. To identify additional studies, we screened all papers in a recent bibliographic paper on CLT research (Adler & Sarstedt, 2021). After excluding duplicates with Soderberg et al., of the 844 papers identified in Adler and Sarstedt (2021), only six contained studies that experimentally examined the influence of psychological distance on a direct measure of construal. One of these, Calderon et al. (2020), was excluded from further screening as it was a direct replication of a prior CLT study. The remaining five papers contained 12 potentially eligible studies. We uncovered an additional seven potentially eligible experiments, from three documents not included in either Soderberg et al. or Adler and Sarstedt: Danziger et al. (2012), Grinfeld et al. (2021), and Liviatan et al. (2008).

The potentially eligible experiments (k = 106) were subjected to a second screening. During this screening, studies were excluded based on seven reasons:

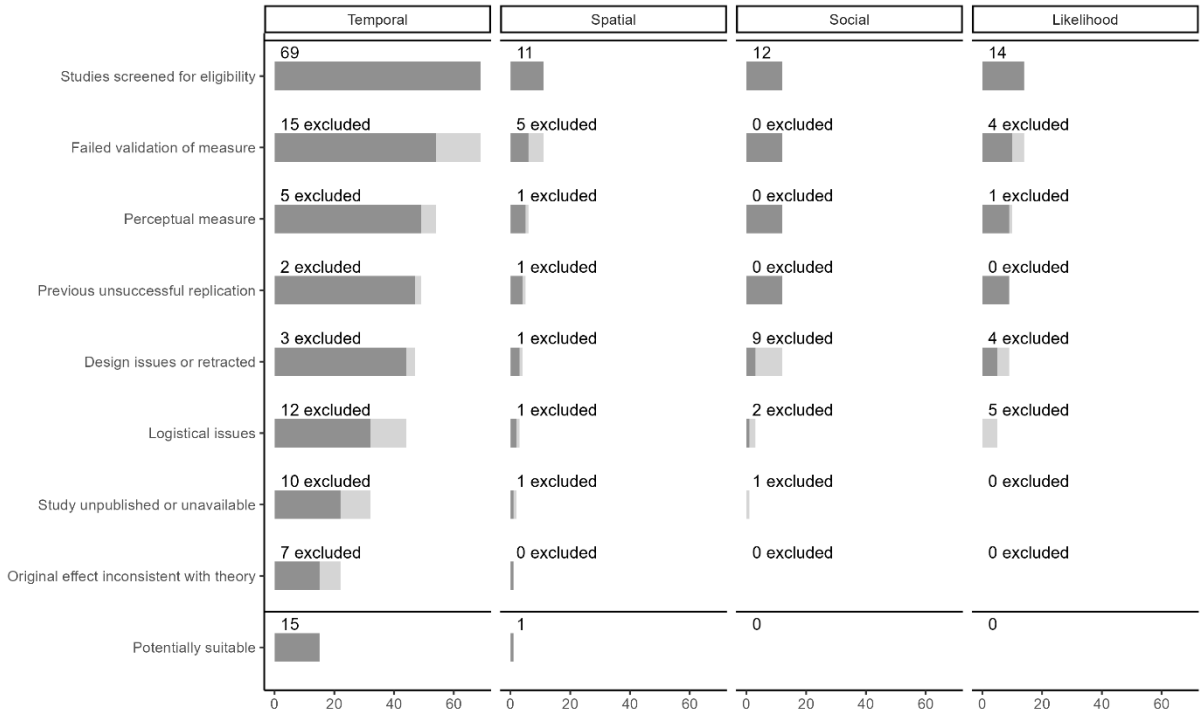
- *Failed validation of measure*. Based on advice from researchers in the field, we pretested several direct measures of mental construal (Mac Giolla et al., 2024). These measures were developed for paper-and-pencil data collections. To validate their use in computerized contexts we examined how the measures responded to direct manipulations of mental abstraction where we directly asked participants to imagine events in more concrete or abstract terms. Out of the five dependent variables we attempted to validate, only one—the Behavior Identification Form (BIF; Vallacher & Wegner, 1989)—worked as intended. Studies using any of the other four dependent measures were excluded.
- *Perceptual measure*. This refers to dependent variables such as the Navon Letters task (Navon, 1977) and the Gestalt Completion Test (Ekstrom et al., 1976). Studies using such measures were excluded on advice from an expert in the field (N. Liberman, personal communication, March 11, 2020).
- *Previous unsuccessful replication*. Studies were excluded if there were previous unsuccessful replication attempts of the studies.
- *Design issues or retracted*. Studies were excluded if there were serious design issues (e.g., the experimental manipulation was confounded with other variables) or if the paper was retracted.
- *Logistical issues*. Studies were excluded if they were methodologically inappropriate for a multi-site project (e.g., were highly culturally specific, required extensive prescreening).
- *Study unpublished or unavailable*. Studies were excluded if they were unpublished or unavailable.

- *Original effect inconsistent with theory*. Studies were excluded if not all of the hypothesized effects were statistically significant ($p < .05$) or if an effect was in the opposite direction from what was predicted.

Figure 1 shows a breakdown of the screening procedure for each of the four types of psychological distance (for the full dataset of coded studies see https://osf.io/x9w4v. For temporal distance, there were several potentially suitable studies. We opted for Liberman and Trope (1998, Study 1) as it is a seminal study in the field that is highly influential (over 2,900 citations on Google Scholar as of December 2022). For spatial distance, we identified only one suitable study, Fujita et al. (2006, Study 1). This is also a highly influential seminal study in the field (over 900 citations on Google Scholar as of December 2022). However, we could not identify any suitable studies for either social distance or likelihood. For this reason, we conducted paradigmatic replications for these two distances. In brief, we extended the basic design of Liberman and Trope (1998) and Fujita et al. (2006), but we replaced the temporal and spatial manipulations with social and likelihood manipulations. For details, see the study-specific protocols below.

Figure 1

Overview of Study Selection Procedure



Note. Light grey bars represent excluded studies.

Labs and Study Participants

Labs were recruited both by efforts of the coordinators of the project (e.g., via a project website, Twitter, Facebook, online forums, and email lists of social psychology networks) and by a call for labs announced by the Association for Psychological Science. The project had no financial resources to pay study participants, and participating labs were therefore free to choose the means of compensation best suited for their local sample (e.g., monetary reimbursement, course credits, voluntary participation). Type of compensation was recorded by each lab (52 labs used course credit, 9 labs used monetary reimbursement, 4 labs used both course credits and monetary reimbursement, 6 labs offered other forms of compensation, and 6

labs gave no compensation to participants). Only individuals 18 years or older were eligible for participation. We set a deadline (January 31, 2024) for labs to confirm that they were willing and able to collect data within a designated period of time.

By the recruitment deadline, 95 labs had signed up to participate. Of these, 78 labs provided data for the analyses. A total of 12,514 participant responses were recorded. We excluded data for three broad reasons:

- Some labs identified cases of participants completing the study more than once, and the additional responses (beyond the first) were removed (n = 241, 1.9% of responses).
- Some labs encountered technical and procedural errors that rendered some data unusable, and these cases were removed (n = 144, 1.2% of responses).
- Some participants did not submit data for the main outcome, and these cases were removed (n = 354, 2.8% of responses).

Details for the repeat participation and technical issues can be found in the supplemental materials (https://osf.io/ptczq). From this point forward, sample sizes reported refer to numbers after these exclusions.

Labs from 27 countries and regions contributed: USA (n = 2,559), Germany (n = 1,296), Turkey (n = 974), Australia (n = 746), the Netherlands (n = 665), China (n = 648), the UK (n = 588), Spain (n = 513), Austria (n = 430), Italy (n = 359), Switzerland (n = 337), Canada (n = 305), Sweden (n = 304), Poland (n = 279), Taiwan (n = 234), Malaysia (n = 200), Israel (n = 198), Slovakia (n = 190), Singapore (n = 148), Denmark (n = 133), France (n = 121), Ireland (n = 117), Serbia (n = 106), New Zealand (n = 104), Belgium (n = 100), Hong Kong (n = 77), and the Philippines (n = 44). Participants were 69.4% (n = 8,173) women, 28.7% (n = 3,374) men, 1.6% (n = 188) non-binary, and 0.3% (n = 41) other genders, with a mean age of 21.6 years (SD = 5.68, Mdn = 20). Data collection started on July 7, 2023 and closed on October 31,

2024. Labs agreed to collect data from at least 100 participants. Of the contributing labs, 12 labs failed to reach 100 participants by the end of the data collection period ($range_n$ = 44–99, M_n = 80), but their data were nonetheless included in the analyses. In total, usable data from 11,775 participants were collected.

Statistical Power

In most situations, statistical power in multi-lab designs is more quickly accrued by increasing the number of labs, rather than increasing the number of participants per lab (Westfall, 2016). To increase the number of contributing labs, we therefore required each lab to collect only a modest number of participants ($n \geq 100$ per lab). This approach means that each individual lab had relatively low power to detect plausible effects. However, across labs, we had ample statistical power to detect relevant effects.

Sample sizes of the four studies ranged from N = 2,926 to N = 2,972. This means that the least powered experiment (i.e., social) had 99.995% power to detect a Hedges' g = 0.23 (i.e., the lower bound of the 95% CI for the bias-corrected meta-analytic effect-size estimate in Soderberg et al., 2015), with 9.11% heterogeneity of effects across labs (the I^2 estimate from the replication with the greatest heterogeneity). The temporal, spatial, and likelihood experiments similarly had 99.995% power to detect this effect. Moreover, the experiments had between 99.991% power (social) and 99.992% (spatial) power to detect what is conventionally considered a small effect, d = .20. The effects detectable with 80% power ranged between d = 0.109 (spatial) and d = 0.110 (social).

Design and Procedure

Each lab received a unique link to the study which was administered via Qualtrics® survey software. For the sake of experimental control, labs were strongly encouraged to collect data in the lab and were asked to consider online data collection only if data collection in the lab proved impossible. In such cases, however, a local sample had to be used (e.g., a university

participant pool, local community members). It was not permissible to use crowdsourcing platforms such as Amazon Mechanical Turk or Prolific. In total, 53 labs (67.9% of all labs) collected data in the lab only, 22 labs (28.2% of all labs) collected data online only, and 3 labs (3.8% of all labs) used a combination of lab and online data collection. This resulted in the collection of data from 7,871 participants in the lab (66.8% of all participants) and 3,715 participants online (31.5%). For one lab that collected data both in-lab and online, a procedural error led to it not being possible to reliably determine which cases were collected in which modality (n = 190, 1.6%). These participants were retained but excluded from the analysis of modality as a moderator. Where possible, online and in-lab data were treated as separate samples for the purposes of analysis, even if the data were collected by the same lab. For one such lab, however, the amount of data collected online was too small to calculate effect sizes. This lab's effects were calculated across the whole sample, and they were excluded from the analysis of modality as a moderator. The effect sizes for the lab for which it was not possible to reliably determine modality were also calculated across the whole sample. Thus, we had a total of k = 79 samples.

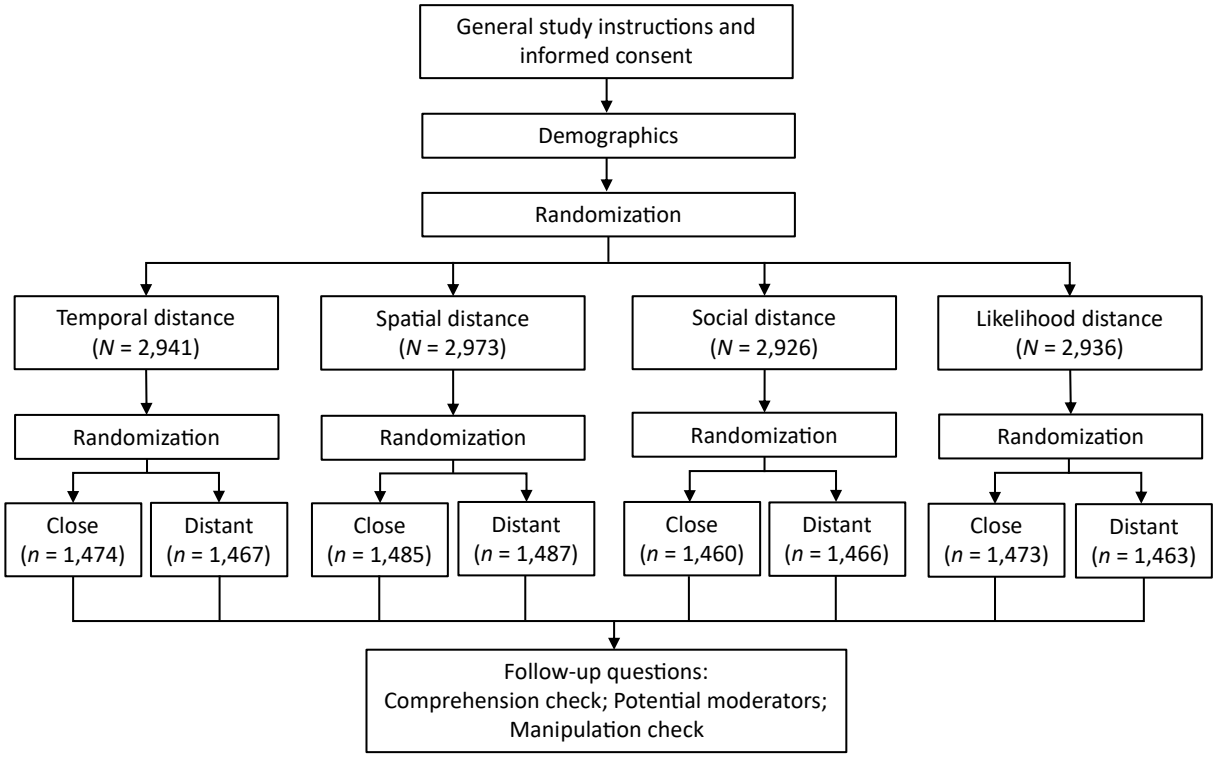
In each lab, participants were randomly assigned to one of four study protocols: temporal distance, spatial distance, social distance, or likelihood. They were then randomly assigned to one of the two experimental conditions in each study (close vs. distant). See Figure 2 for a flowchart of the full procedure. For ethical or practical reasons, some labs required minor procedural modifications (e.g., omission of recording ethnicity data), and these modifications are noted in the supplemental material (https://osf.io/ptczq).

Durations for the experiments were automatically recorded by the survey platform. Some durations recorded were clearly incorrect (e.g., multiple days), likely due to technical issues (e.g., the platform treating a completed survey as though it had not been submitted). Examining the raw data from responses that took more than 60 minutes (n = 82, 0.7% of the

sample), there were no obvious defects or unusual behavior. These data were retained for analysis, but they were excluded from calculating typical durations of the study. The median time required for participants to complete the study was 7.69 minutes (range: 1.03 – 59.70; $Mdn_{\text{Liberman \& Trope., (1998, Study 1)}} = 7.85$; $Mdn_{\text{Fujita et al., (2006, Study 1)}} = 7.36$; $Mdn_{\text{Social}} = 8.08$; $Mdn_{\text{Likelihood}} = 7.45$).

Figure 2

Flowchart of the Study Procedure



Note. Sample and group sizes are reported after the removal of participants with incomplete data.

General Instructions

Prior to providing informed consent, all participants were informed about approximately how long the study would take, roughly what it would consist of, and what, if any, compensation

they would receive for participating. Participants were also informed that participation was voluntary; that they could withdraw at any stage without any explanation needed; that their responses were anonymous, insofar as answers could not be traced to any individual; and that the anonymous data would be made openly available to other researchers. The exact formulation of labs' consent forms varied due to differences in local IRB requirements. Labs' verbatim consent forms can be found on OSF (https://osf.io/zywms/). Participants were required to actively check a box on the computer screen to indicate that they had understood the information and provided their consent. If participants consented, they were taken to the next page that enquired about demographic information.

For demographics, participants were asked the following questions with answer options provided in brackets: age (numeric entry in years), gender (male, female, non-binary, other), nationality (list of countries), ethnicity (free text), occupation (employed, student, other), and highest education level achieved (primary school, secondary level [high-school], college/university, post-graduate). If college/university or post-graduate was selected, an additional question about primary subject area was asked. Enquiring about demographics prior to the experiment was required due to the experimental manipulation in the social distance study (see study protocols below).

Participants were then randomly assigned to one of the four study protocols. The "evenly present elements" option in Qualtrics was applied to ensure that the randomization process produced approximately equal group sizes within each lab.

Protocols

We kept the two direct replication protocols—temporal distance (Liberman & Trope, 1998) and spatial distance (Fujita et al., 2006)—as similar as possible to the original studies, but made some necessary adjustments. Primarily, changes concerned making the protocols

appropriate for an international data collection instead of a region-specific one and switching from paper-and-pencil data collection to collecting data via an electronic questionnaire. Below we present a brief description of the four study protocols. The full study protocols for all four studies, as well as a detailed description of the differences and similarities between the original and replication experiments are available on OSF (https://osf.io/zywms/).

The translation of the study materials followed the procedure used by Jones et al. (2021) in a recent multi-lab replication project (for the original translation procedure, see Brislin, 1970). Labs conducting the study in a language other than English were asked to coordinate the translation of the protocols to their own language, and translations were then independently back-translated to ensure accuracy (see OSF for full details on the translation procedure, https://osf.io/awzfc).

To maximize transparency, each lab was asked to make a video recording of their procedure for administering the study in the lab using a mock participant. The videos are available on OSF (https://osf.io/r89ks/).

1. The Effect of Temporal Distance on the Behavior Identification Form (Liberman & Trope, 1998, Study 1).

Original Study. In the original study, participants ($N = 32$) were asked to complete an amended version of Vallacher and Wegner's (1989) Behavior Identification Form (BIF).[1] The full scale consists of 25 activities (e.g., "locking a door"), but the original study excluded six items which were deemed as not being a good fit for the specific sample used. These were "joining the army", "picking an apple", "chopping down a tree", "voting", "climbing a tree" and "growing a garden" (for a list of the full 25 items of the BIF, see Appendix A). The remaining 19 items were shown to participants who were asked to choose one of two alternative

[1] The original study also had a first part where participants described seven activities in their own words. This writing task was not included in the replication study because (1) reviewers advised against including it, and (2) because the data would be infeasible to analyze within a multi-lab approach.

descriptions of the activity: one relatively concrete description (e.g., "putting a key in the lock") and one relatively abstract description (e.g., "securing the house"). Temporal distance was manipulated by telling participants to imagine engaging in the activities either "tomorrow" (close condition) or "next year" (distant condition). The study was conducted in the lab using paper and pencil. An independent *t*-test of the BIF scores (i.e., the number of abstract descriptions chosen) revealed that participants chose more abstract activity descriptions in the distant than in the close condition, $d = 0.92$, 95% CI [0.18, 1.66]. The effect size reported here was calculated based on the statistics reported in the paper.

Replication. We received the original study materials from the authors and followed them as closely as possible. Participants were first presented with standard instructions for filling out the BIF. As with the original study, this was closely based on the instructions for Vallacher and Wegner's (1989) original scale. To manipulate temporal distance, participants were asked to imagine engaging in the 19 behaviors of the abridged BIF either "next year" (temporally distant) or "tomorrow" (temporally close). In line with the original instructions, each BIF item was also phrased in accordance with the experimental condition. For example, participants were asked to "Think about yourself painting a room [next year]/[tomorrow]".[2]

In addition, for the replication study we included the six BIF items that were excluded in the original study, for exploratory purposes. The original authors excluded the items because they were deemed irrelevant for their sample. Our replication study, however, used a more heterogeneous sample than the original study and these items could provide interesting information about effects on all 25 activities. The six added items were presented separately on a subsequent page, so they could not affect the responses on the 19 items included in the original study (for the 19 original items used in the primary analyses, see Appendix A).

[2] In the English version of the materials, we changed the original formulation "the next year" to "next year" because we believed the revised wording sounded more natural.

2. The Effect of Spatial Distance on the Behavior Identification Form (Fujita et al., 2006, Study 1).

Original Study. This study used the same basic approach as Liberman and Trope (1998, Study 1), except that spatial distance was manipulated instead of temporal distance. We obtained a copy of the original materials from the authors. In the study, participants ($N = 68$)—students at a university in New York City—were asked to imagine a scenario in which they had moved to a new apartment. In the spatially close condition, participants were told the apartment was located "outside of New York City, which is just under 3 miles away." In the spatially distant condition, they were told the apartment was "just outside of Los Angeles, which is over 3,000 miles away."[3] Participants were then asked to imagine performing a number of activities in their new apartment. Specifically, they were asked to imagine performing 13 different actions. These actions were taken from the BIF (the remaining 12 items on the BIF were deemed irrelevant to the scenario). See Appendix A for the 13 items included in the original study. Participants then selected their preferred description of the 13 actions. The study was conducted in the lab using paper and pencil. An independent *t*-test of the summed BIF-scores revealed that participants chose more abstract activity descriptions in the distant than in the close condition, $d = 0.55$, 95% CI [0.06, 1.03]. The effect size reported here was calculated based on the statistics reported in the paper.

Replication. Except for two necessary adjustments, the instructions in the replication were identical to the original instructions. First, adjustments were made to account for each

[3] In the published article, it is reported that participants were to "imagine a scenario in which they were helping a friend move into a new apartment" and that the apartment is either "outside of New York City, about 3 miles away from here" and "outside of Los Angeles, about 3,000 miles away from here" (Fujita et al., 2006, p. 279). We obtained the original materials from the authors, and these materials contain different instructions from the published version. The original authors confirmed that the version they sent to us was the version in fact used in the original study. We have based our study on the original materials obtained from the authors, not on the description in the article.

lab's specific location. Labs were instructed to use their home city for the close condition, and a distant city in their own country for the distant condition. Furthermore, for labs in countries and regions using the metric system, the distances were expressed in kilometers rather than miles. For example, when data collection took place in the city of Gothenburg, Sweden, the instructions for the close/far conditions were "[just outside of Gothenburg, which is under 5 km away]/[just outside of Kiruna, which is over 1,200 km away]." Second, participants were asked to "choose" rather than "circle" their preferred items on the BIF. This was because the replication was conducted on a computer rather than using paper and pencil. All other instructions were identical.

The original authors excluded 12 items from the BIF because they were deemed irrelevant to the scenario. For exploratory purposes, these 12 items were presented separately on a subsequent page, so they could not affect the responses on the 13 original items that were included in the primary analyses.

3. The Effect of Social Distance on the Behavior Identification Form. There was no suitable original study examining the effect of social distance on the BIF. Instead, we conducted a paradigmatic replication where we adapted the design of Liberman and Trope (1998, Study 1; used in the temporal-distance replication above) to test the effect of social distance. Participants were given the same basic instructions on how to fill out the BIF as in Liberman and Trope's study. However, instead of manipulating temporal distance, we administered a social distance manipulation inspired by Yan et al. (2016, Experiment 3). Participants were asked to imagine a target person that was either similar (close target) or dissimilar (distant target) to themselves in terms of age, gender, educational background, and personal interests.

The description of the target was modeled on the demographic information that participants provided at the start of the study. In the socially close condition, the target's age and gender matched that of the participant. Age was calculated by adding two years to the

participant's own reported age. The target's name was drawn randomly from a list of six common first names, specific to the country of data collection, of men (for male participants) or women (for female participants) born in the 1960s (for older participants) and 1990s (for younger participants). For participants who reported being non-binary or of other gender, the name for socially close targets was selected randomly from a list of six common gender-neutral names specific to the country of data collection.

For participants in the socially distant condition, the target's age and gender did not match that of the participant. For participants below 40 years, 20 years were added to the reported age. For participants 40 years or older, 20 years were subtracted from the reported age. The target's name was again drawn randomly from a list of six common first names, specific to the country of data collection. However, for male participants the target name was a common name of women born in the 1960s (for younger participants) and 1990s (for older participants). For female participants the target name was a common name of men born in the 1960s (for younger participants) and 1990s (for older participants). For participants who reported being non-binary or of other gender, the name was randomly chosen from all of the 12 male or female names born in the 1960s (for younger participants) and 1990s (for older participants). In a pretest ($N = 300$), this manipulation produced a very large effect on ratings of perceived social distance, $d = 3.15$, 95% CI [2.81, 3.49] (for details, see https://osf.io/ahyvj/). Below are example descriptions of a socially close and a socially distant target for a female participant, 24 years old, from the UK:

> *(Socially close)* Hannah is a woman aged 26. Hannah has an educational background that is similar to yours and she shares several of your personal interests. In other words, Hannah is a person with whom you have a lot in common.

(Socially distant) Paul is a man aged 44. Paul has an educational background that is very different from yours and he does not share any of your personal interests. In other words, Paul is a person with whom you have little in common.

When filling out the BIF, participants were asked to imagine that the target was performing the activities. Using the socially close example above, the instructions would read: "For each behavior in the list, you will be asked to imagine Hannah performing them." Again, each BIF item was phrased in accordance with the experimental condition. For example, participants were asked to "Think about Hannah locking a door." For this study, participants filled out the full 25-item BIF.

4. The Effect of Likelihood on the Behavior Identification Form. There was no suitable original study examining the effect of likelihood on the BIF. Instead, we conducted a paradigmatic replication where we adapted the design of Fujita et al. (2006, Study 1), using an experimental manipulation of likelihood developed by Wakslak et al. (2006, Study 1). Participants were asked to imagine that they had been asked to help a friend move. Participants were then told that the friend would only be moving if they were offered a job they had applied for. In the high likelihood condition, participants were told the friend thought there was a 95% chance they would get the job. In the low likelihood condition, they were told that the friend thought there was a 5% chance they would get the job. Participants were then asked to fill out an abridged nine-item BIF and to imagine performing the activities in relation to the scenario of helping a friend move. The nine BIF items were selected based on pretesting. In the pretest, an online sample (N = 183) rated the relevance of the 25 BIF items to the scenario of helping a friend move to a new apartment. Ratings were made on a five-point scale (-2 = *very irrelevant*, -1 = s*omewhat irrelevant*, 0 = *neither relevant nor irrelevant*, 1 = s*omewhat relevant*, 2 = *very relevant*). Based on a pre-registered analysis plan, items with a mean rating significantly greater than 0 were selected as relevant for the scenario (for study details as well as the pre-registration,

see https://osf.io/h7f4q/). For the nine relevant items, see Appendix A. For exploratory purposes, the remaining 16 BIF items were presented on a subsequent page, so that they could not affect the responses on the nine items that were included in the primary analyses.

Follow-Up Questions

In addition to the primary outcome measures of the four studies, we also included several follow-up questions. These were presented to all participants and included a comprehension check, an assessment of participants' mood, a dispositional measure of analytic versus holistic thinking, and a manipulation check. The follow-up questions were included for exploratory analyses and robustness checks.

Comprehension Check

Participants were asked a multiple-choice question about the distance-manipulation instruction that they had received at the beginning of the study. Specifically, they were asked *how* they had been asked to imagine the activities that they had previously rated. They were then given six response options, presented in random order, one of which was correct. The exact formulation of the question and the six response options were tailored to the specific study. For example, the question for the temporal study read: "You were previously asked to imagine engaging in a series of activities (e.g., making a list, painting a room, brushing teeth). When were these events supposed to take place?" (1 = *next year*, 2 = *tomorrow*, 3 = *in five weeks*, 4 = *in six months*, 5 = *18 months from now*, 6 = *in two years*). All comprehension checks are presented in Appendix B.

Self-Rated Mood

Several previous CLT studies have examined mood as a potential confound to the effect of psychological distance on construal level. For example, Wakslak et al. (2006, Study 1) measured participants' mood to check that the experimental groups did not differ on this variable. This is a legitimate concern, given that positive mood has been shown to positively

correlate with mental abstraction (Fredrickson & Branigan, 2005). To examine mood as a potential confound, self-ratings of participants' mood were collected in all four study protocols using the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). The scale consists of 20 items measuring different affective states (e.g., interest, stress, excitement), which are rated using a five-point scale (1 = *very slightly or not at all*, 5 = *extremely*).

Tendency for Analytic versus Holistic Thinking

Participants completed the 12-item Analysis-Holism Scale (AHS; Martín-Fernández et al., 2022). The AHS measures analytic–holistic thinking style on four subdomains: causality, attitude towards contradictions, perception of change, and locus of attention. To measure participants' locus of attention, for example, participants are asked to rate the degree to which they agree with statements such as "The whole, rather than its parts, should be considered in order to understand a phenomenon" and "It is more important to pay attention to the whole context rather than the details". Ratings were made on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*).

Manipulation Check

At the end of the survey, participants were randomly assigned to validate the experimental manipulation from one of the three studies in which they had not taken part. Thus, participants were presented with a manipulation check for a different psychological distance than that to which they had already been exposed. This allowed us to gauge the strength of each manipulation in a way that did not rely on participants' retrospective memory for the manipulation and was minimally affected by their previous responses.

Participants received a brief description of the type of psychological distance examined in that study and were then presented with the experimental manipulation from the close or distant condition, phrased as similarly as possible to the manipulation in the actual study. For example, the manipulation check for the close condition in the temporal study read: "Events

can feel closer or more distant in time from the present moment. When events feel like they will happen soon, they are said to be temporally close. In contrast, when events feel like they will not happen for a very long time, they are said to be temporally distant. To what extent does something taking place tomorrow feel temporally close or distant?" Participants then provided their response on a seven-point scale ranging from 1 (*very close*) to 7 (*very distant*). For the exact wording of each manipulation check, see Appendix C.

Assessment of Statistical Power

We conducted three sets of power analyses to provide an evaluation of the current replications in relation to the existing literature examining the effects of psychological distance on construal level. Previous experiments examining these effects constitute the body of evidence substantiating the conceptual hypotheses we aimed to test with the present replications. As such, this literature provides a useful point of reference for interpreting the present results.

First, we calculated the sample sizes needed to detect the replication effects across different levels of power (1% to 99%). This analysis provides information that can be applied when planning future studies. Additionally, it provides an intuitive measure of how "visible" the estimated effect is (viz. how many people would need to be observed to reliably detect the effect).

Second, we calculated the effect sizes that the extant literature (i.e., experiments examining direct effects of psychological distance on construal level) was sensitive to detect at 80% power. To calculate these effect sizes, we extracted the sample sizes and number of groups in the design from each of the experiments we screened for replication (see Figure 1). We then plotted a frequency distribution of these effect sizes with the replication summary effects, and for each replication, we calculated the percentage of experiments in the past literature that had 80% power to detect an effect at least as large as the summary effect size for the replication.

This comparison provides information about how the effects estimated by the replications compare to the sensitivity of previous experiments. The median effect size for which previous experiments (k = 100) had 80% power was d = 0.68, with a range from d = 0.18 to d = 1.32. Assuming researchers implicitly or explicitly reasoned about the plausible size of the effects they were studying when determining their sample sizes, this power analysis can offer a comparison of the replication results to benchmarks implied by the literature.

Finally, we calculated the power that previous experiments had to detect the summary effect sizes from the four present replications. These power estimates were based on the group sizes extracted from the previous studies. Assuming that the replications provide reasonable estimates of direct effects of psychological distance on mental abstraction, examining these values provides information about how well-powered prior experiments have been to detect effects of interest.

Valence Differences Between Response Options for the BIF Items

A reviewer of the Stage 1 Report commented that the response options for the items on the BIF may be systematically biased such that the abstract options tend to be more positively valenced than the concrete options. The reviewer pointed out that this bias might be particularly problematic for the social distance replication, wherein participants might be motivated to provide more positive descriptions for socially closer targets. To address this issue, we recruited participants on Prolific (prolific.com) to rate the valence of the response options independently (i.e., each option was rated on its own; N = 300) and comparatively (i.e., the two options for each BIF item were compared in terms of valence; N = 302). The results of these pretests indicated that abstract options were indeed rated as more positive than concrete options, $d_{\text{independent}}$ = 0.67, 95% CI [0.64, 0.70] and $d_{\text{comparative}}$ = 0.75, 95% CI [0.72, 0.77].

To assess the plausibility of these valence differences as a threat to validity, we joined our pretest data with data from a previous social distance experiment with a design similar to

ours. We requested the data for Yan et al. (2016, Experiment 3), which the authors graciously provided, and analyses with these data suggested that although the valence differences (rated comparatively, measured as a standardized mean difference for each item) predicted responses on BIF items, controlling for the valence differences did not change the effect of social distance, and there was no significant interaction between valence and the distance manipulation. As such, prior to conducting the present studies, we believed the valence differences posed little or no threat to the validity of the replications. However, as a robustness check, we conducted analyses for each replication testing whether participants' responses to the BIF were influenced by the valence differences and whether the valence differences interacted with the distance manipulations. These analyses used the measurements from the comparative pretest ratings. For details on the pretests and the analyses of previous data, see https://osf.io/g6d5v.

Results

Project Compendium

Materials for this project are available in a compendium comprising two digital repositories, located on the Open Science Framework (https://osf.io/ra3dp/) and GitHub (https://github.com/RabbitSnore/CLIMR). Raw data, which include all variables to reproduce the analyses as well as additional exploratory variables, are available on OSF. Detailed analysis reports and supplemental information are archived on OSF. These reports are available with embedded graphics on GitHub. The R code (R Core Team, 2022) for performing the main analyses was written and registered prior to data collection. The code was only altered to correct errors, facilitate importation and formatting of raw data, and/or to troubleshoot technical issues. The most up-to-date version of the code is available on GitHub, and the version of the code repository as it existed prior to data collection is archived on OSF. Instructions for reproducing the analyses and data visualizations are provided in the readme file on GitHub.

Analytic Strategy

For each experiment from each lab, we calculated an effect size for the primary comparison of interest. Three labs contributed both in-lab and online samples. For one of these labs, effect sizes were calculated for the in-lab and online samples separately. For another one of these labs, there were too few cases in the online data set, so one effect size was calculated for both data sources (and this lab's data was excluded from the analysis of modality as a moderator). For the third lab, an error prevented identification of each participant's modality, so effect sizes were calculated for the whole sample. Thus, with 78 contributing labs, we had 79 effect sizes for each experiment, except for the likelihood replication, for which we had 78 effect sizes because a technical issue with one lab rendered the data from this experiment unusable. In all four experiments, the critical comparison was between psychologically close and psychologically distant conditions. Because our dependent variable (the BIF) uses sum scores, we calculated standardized mean differences (*d*) as effect sizes.

For each experiment, we conducted a random effects meta-analysis of the effect sizes from each contributing lab. We used the *metafor* package (Viechtbauer, 2010) to compute meta-analytic statistics for our three main analyses. First, we provide an assessment of the presence of an effect by assessing whether 0 is excluded by the lower bound of the 95% confidence interval for the meta-analytic estimate. Second, to assist with substantive interpretation, we provide unstandardized effect estimates in the scale of the dependent variables. Third, we provide an assessment of the heterogeneity across contributing laboratories.

Additionally, we report the number of participants that would be required to achieve 80% and 95% power to detect the estimated effect for each replication, assuming a two-group experimental design. We report the percentage of previous experiments that had the sensitivity to detect the estimated effect of each replication at 80% power or higher. We also report the median statistical power for each replication effect size for sample sizes from previous experiments.

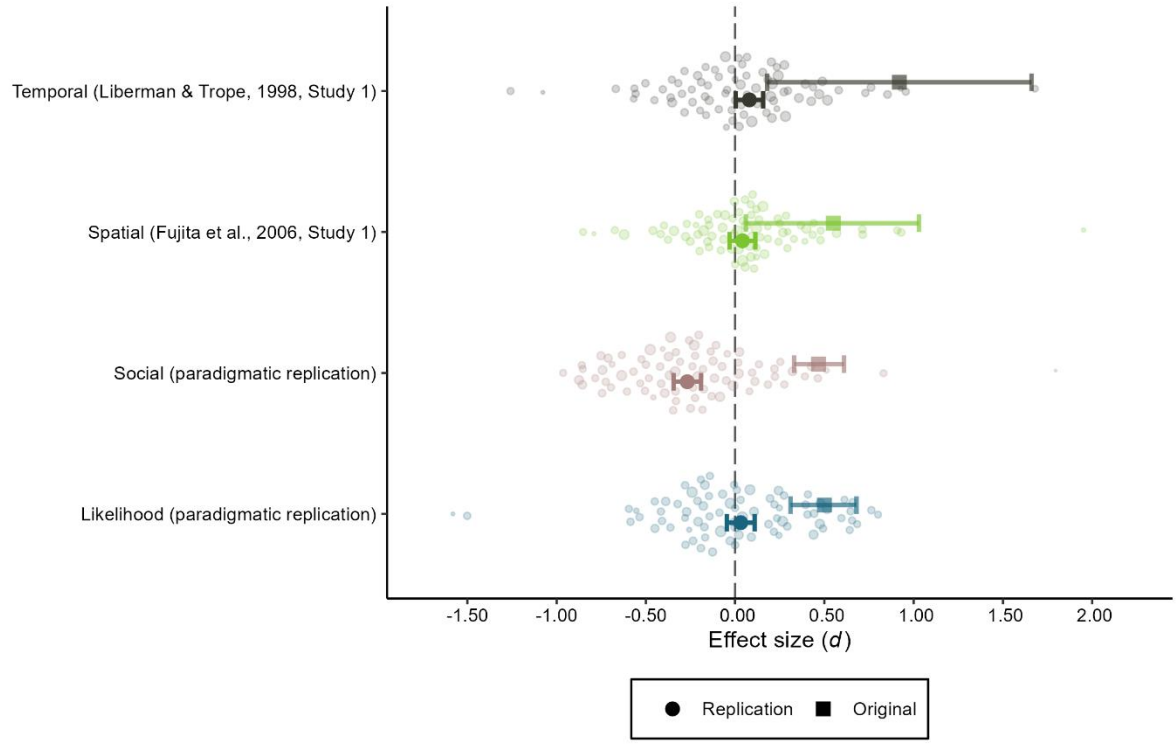
For each replication experiment, we also provide an estimate of the average effect of the manipulation on the manipulation check. This estimate and its corresponding 95% CI are derived from a random effects meta-analysis.

Replication Effects

Figure 3 displays the results for each of the four experiments. Forest plots for the four individual study protocols can be found at https://osf.io/c37dw. Figure 4 displays the power analyses conducted to contextualize the replication results in relation to the existing literature.

Figure 3

Replication and Original Effect Sizes for Each Experiment



Note. Individual points represent replication effect size estimates from contributing labs. Symbols with error bars represent the meta-analytic effect size from the replications (dots) and the original effect sizes (squares), with 95% confidence intervals as error bars. For social distance and likelihood, "original" effect sizes are the meta-analytic estimates for those distances from Soderberg et al. (2015).

Temporal Distance

The temporal distance studies—direct replications of Liberman and Trope (1998, Study 1)—yielded a meta-analytic effect of d = 0.08, 95% CI [0.003, 0.16], 95% prediction interval [-0.13, 0.29]. The lower bound of the 95% confidence interval for this estimate excluded zero. Transforming the meta-analytic estimate to an unstandardized effect, we found that on average participants gave 0.32, 95% CI [0.01, 0.63], more abstract responses on the study specific 19-item BIF when they were assigned to the distant condition, compared to the close condition. Across labs, participants' mean BIF scores were 9.09 (SD = 3.92) in the close condition and 9.41 (SD = 4.05) in the distant condition. A Q-test indicated an amount of heterogeneity not greater than what would be expected by random sampling error, $Q(78) = 93.88$, $p = .106$, $I^2 = 8.07\%$, $\tau = 0.098$.

In a two-group experiment, the meta-analytic effect size for the temporal studies would require N = 4,927 participants to detect with 80% power and N = 8,155 participants to detect with 95% power. Of the experiments in the previous literature, none had 80% or more power to detect the meta-analytic effect size for this replication. Across sample sizes from previous experiments, the median power for this effect size estimate was 5.1%.

Across labs, the meta-analytic estimate for the effect on the manipulation check for temporal distance was d = 0.98, 95% CI [0.88, 1.07]. The lower bound of the 95% confidence interval for this estimate excluded zero.

Spatial Distance

The spatial distance studies—direct replications of Fujita et al. (2006, Study 1)—yielded a meta-analytic effect of d = 0.04, 95% CI [-0.03, 0.11], 95% prediction interval [-0.03, 0.11]. The lower bound of the 95% confidence interval for this estimate did not exclude zero. Transforming the meta-analytic estimate to an unstandardized effect, we found that on average participants gave 0.11, 95% CI [-0.09, 0.31], more abstract responses on the study-specific 13-

item BIF when they were assigned to the distant condition, compared to the close condition. Across labs, participants' mean number of abstract choices on the BIF were 8.31 (*SD* = 2.72) in the close condition and 8.41 (*SD* = 2.68) in the distant condition. A *Q*-test indicated an amount of heterogeneity not greater than what would be expected by random sampling error, $Q(78) = 72.20$, $p = .664$, $I^2 = 0.01\%$, $\tau = 0.003$.

In a two-group experiment, the meta-analytic effect size for the spatial studies would require $N = 18{,}682$ participants to detect with 80% power and $N = 30{,}929$ participants to detect with 95% power. Of the experiments in the previous literature, none had 80% or more power to detect the meta-analytic effect size for this replication. Across sample sizes from previous experiments, the median power for this effect size estimate was 3.7%.

Across labs, the meta-analytic estimate for the effect on the manipulation check for spatial distance was $d = 1.22$, 95% CI [1.10, 1.35]. The lower bound of the 95% confidence interval for this estimate excluded zero.

Social Distance

The social distance studies yielded a meta-analytic effect of $d = -0.27$, 95% CI [-0.35, -0.19], 95% prediction interval [-0.46, -0.08]. The lower bound of the 95% confidence interval for this estimate did not exclude zero, but the upper bound did, suggesting an effect in the direction opposite the hypothesis. Transforming the meta-analytic estimate to an unstandardized effect, we found that on average participants gave 1.39, 95% CI [-1.79, -0.99] fewer abstract responses on the full 25-item BIF when they were assigned to the distant condition, compared to the close condition. Across labs, participants' mean number of abstract choices on the BIF were 10.60 (*SD* = 5.39) in the close condition and 9.28 (*SD* = 4.85) in the distant condition. A *Q*-test indicated an amount of heterogeneity not greater than what would be expected by random sampling error, $Q(78) = 85.41$, $p = .265$, $I^2 = 6.77\%$, $\tau = 0.090$.

In a two-group experiment, the meta-analytic effect size for the social studies would require $N = 439$ participants to detect with 80% power and $N = 726$ participants to detect with 95% power. Of the experiments in the previous literature, 2% ($k = 2$) had at least 80% power to detect the meta-analytic effect size for this replication. Across sample sizes from previous experiments, the median power for this effect size estimate was 19.4%.

Across labs, the meta-analytic estimate for the effect on the manipulation check for social distance was $d = 1.58$, 95% CI [1.45, 1.71]. The lower bound of the 95% confidence interval for this estimate excluded zero.

Likelihood

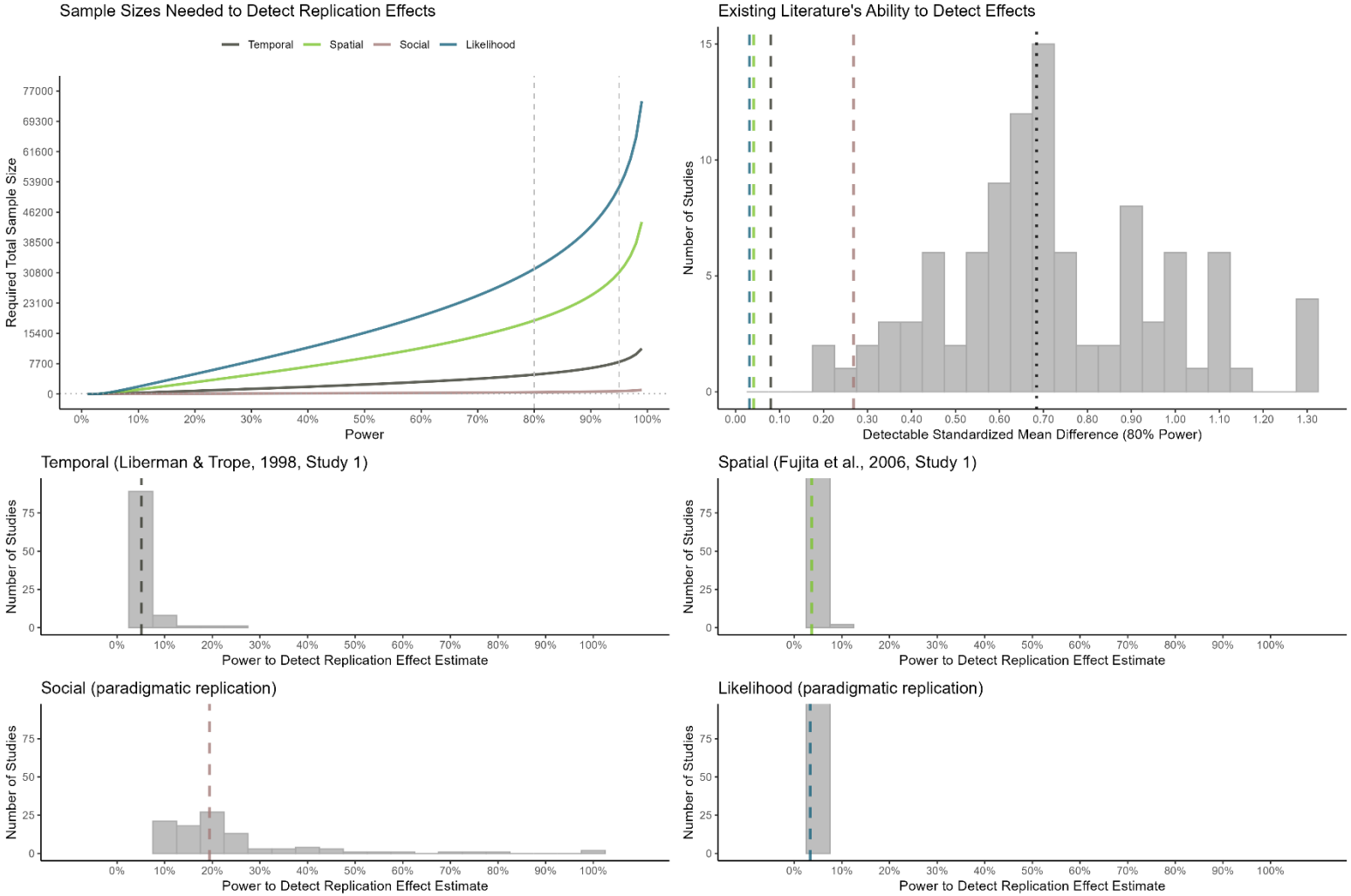
The likelihood studies yielded a meta-analytic effect of $d = 0.03$, 95% CI [-0.05, 0.11], 95% prediction interval [-0.18, 0.25]. The lower bound of the 95% confidence interval for this estimate did not exclude zero. Transforming the meta-analytic estimate to an unstandardized effect, we found that on average participants gave 0.06, 95% CI [-0.09, 0.22] more abstract responses on the study specific 9-item BIF when they were assigned to the distant condition, compared to the close condition. Across labs, participants' mean number of abstract choices on the BIF were 4.53 ($SD = 2.04$) in the close condition and 4.62 ($SD = 2.01$) in the distant condition. A Q-test indicated an amount of heterogeneity not greater than what would be expected by random sampling error, $Q(77) = 92.05$, $p = .116$, $I^2 = 9.11\%$, $\tau = 0.104$.

In a two-group experiment, the meta-analytic effect size for the likelihood studies would require $N = 31{,}787$ participants to detect with 80% power and $N = 52{,}626$ participants to detect with 95% power. Of the experiments in the previous literature, none had 80% or more power to detect the meta-analytic effect size for this replication. Across sample sizes from previous experiments, the median power for this effect size estimate was 3.4%.

Across labs, the meta-analytic estimate for the effect on the manipulation check for likelihood was d = 1.98, 95% CI [1.81, 2.14]. The lower bound of the 95% confidence interval for this estimate excluded zero.

Figure 4

Evaluation of the Statistical Power of the Existing Literature



Note. In the top left panel, each curve represents the relationship between sample size (total *N* for a two-group experiment) and statistical power for the meta-analytic effect size estimates for each replication. Vertical lines are drawn at 80% and 95% power, and a dotted horizontal line is drawn at the median sample size ($N = 70$) in the literature examining direct effects of psychological distance. The top right panel displays the frequency distribution for effect sizes for which previous studies ($k = 100$) had 80% power, based on their group size and design. Colored vertical dashed lines are drawn at the meta-analytic effect size estimate for each replication, and a dotted line is drawn for the median effect size for which the existing literature has 80% power ($d = 0.68$). The panels on the lower half of the figure display the frequency distribution of statistical power previous studies had (based on their group size and design) to detect each of the four meta-analytic effect size estimates from the replications. Vertical dashed lines are drawn at the median power to detect the replication effect size.

Robustness Check: Comprehension Check Failures

For each experiment, we asked questions to check whether participants understood the stimuli. As a robustness check, we excluded observations for which participants failed to respond correctly. Across all studies, 2,044 (17.4%) people out of 11,775 failed the comprehension check. For in-lab data collections, 1,276 (16.2%) out of 7,886 people failed the comprehension check. For online data collections, 768 (19.7%) out of 3,889 people failed the comprehension check.

In the temporal replications, 632 (21.5%) out of 2,941 participants failed the comprehension check. Excluding comprehension check failures, the temporal distance studies yielded a meta-analytic effect of similar magnitude, but the confidence intervals now included zero, $d = 0.08$, 95% CI [-0.004, 0.17], 95% prediction interval [-0.14, 0.30]. Transforming the meta-analytic estimate to an unstandardized effect, we found that on average participants gave 0.32, 95% CI [-0.02, 0.68], more abstract responses on the study specific 19-item BIF when they were assigned to the distant condition, compared to the close condition. A Q-test indicated an amount of heterogeneity not greater than what would be expected by random sampling error, $Q(78) = 93.96$, $p = .105$, $I^2 = 1.07\%$, $\tau = 0.10$.

Excluding comprehension check failures had little influence on the effect size estimates for the spatial, social, and likelihood replications. Analyses with these data excluded are available in the supplemental materials (https://osf.io/z5axm).

Robustness Check: Valence Differences in Response Options for BIF Items

To examine the potential influence on the results of the valence differences in the response options for the items on the BIF, we fit a series of mixed-effects logistic regression models for each experiment. The first model predicted responses on BIF items (0 = concrete, 1 = abstract) using the distance manipulation as a fixed effect, with random intercepts for each

participant nested in each lab and random intercepts for each item. The second model added the standardized mean difference (*d*) in valence for the response options for each item as a fixed effect. These valence differences were taken from our pretest of the BIF (https://osf.io/g6d5v/), range d = -0.05 to d = 1.51. The third model added the interaction term between the distance manipulation and the valence differences. We compared these models with likelihood ratio tests, to identify a model for retention and interpretation. Models that offered significant improvement ($p < .05$) were preferred over previous models.

In a model including the valence differences, the coefficient for the distance manipulation is interpretable as the estimated effect (in log odds scale), at the average value of the response option valence differences. In a model including the interaction term, the coefficient for the distance manipulation is interpretable as the estimated effect when the valence difference is zero. The valence difference coefficient is interpretable as the extent to which participants preferred to select the abstract option as that option was more positively valenced. The interaction term is interpretable as the extent to which the manipulation's effect is amplified (if positive) or mitigated (if negative) as the valence difference increases.

Figure 5 displays the predicted probability of selecting the more abstract BIF response option, as a function of valence differences, for each experiment. These predicted probabilities were calculated from the retained models (see below).

Temporal Distance

For the temporal distance replications, likelihood ratio tests indicated that adding valence differences to the model offered significant improvement to the model, $\chi^2(1) = 16.79$, $p < .001$, and adding the interaction between the manipulation and valence differences offered further significant improvement to the model, $\chi^2(1) = 8.86$, $p = .003$. In the retained model, the coefficient for the manipulation was $b = 0.22$, 95% CI [0.11, 0.34]. The coefficient for valence

differences was $b = 1.38$, 95% CI [0.88, 1.88]. The coefficient for the interaction was $b = -0.15$, 95% CI [-0.25, -0.05].

To summarize, participants tended to select the more abstract option more frequently when it was more positive, and when the abstract option was more positive, the effect of the temporal distance manipulation was mitigated. When there was no difference in valence between the response options, the lower bound of the 95% confidence interval for the estimated effect of the temporal distance manipulation excluded zero. Converting from log odds to *d*, the estimated effect for the temporal distance manipulation accounting for valence differences was $d = 0.12$, 95% CI [0.06, 0.19].

Spatial Distance

For the spatial distance replications, likelihood ratio tests indicated that adding valence differences to the model offered significant improvement to the model, $\chi^2(1) = 9.21$, $p = .002$, and adding the interaction between the manipulation and valence differences did not offer significant improvement to the model, $\chi^2(1) = 0.53$, $p = .465$. In the retained model, the coefficient for the manipulation was $b = 0.04$, 95% CI [-0.03, 0.12]. The coefficient for valence differences was $b = 1.42$, 95% CI [0.66, 2.17].

To summarize, participants tended to select the more abstract option more frequently when it was more positive. At the average level of difference in valence between the response options ($d = 0.92$), the lower bound of the 95% confidence interval for the estimated effect of the spatial distance manipulation did not exclude zero. Converting from log odds to *d*, the estimated effect for the spatial distance manipulation accounting for valence differences was $d = 0.03$, 95% CI [-0.02, 0.06].

Social Distance

For the social distance replications, likelihood ratio tests indicated that adding valence differences to the model offered significant improvement to the model, $\chi^2(1) = 16.16$, $p < .001$,

and adding the interaction between the manipulation and valence differences offered further significant improvement to the model, $\chi^2(1) = 49.07$, $p < .001$. In the retained model, the coefficient for the manipulation was $b = 0.01$, 95% CI [-0.09, 0.12]. The coefficient for valence differences was $b = 1.04$, 95% CI [0.68, 1.41]. The coefficient for the interaction was $b = -0.34$, 95% CI [-0.43, -0.24].

To summarize, participants tended to select the more abstract option more frequently when it was more positive, and when the abstract option was more positive, the effect of the social distance manipulation was more negative. In other words, when the target person was socially close, participants selected the abstract option more frequently, when that option was more positive. When there was no difference in valence between the response options, the lower bound of the 95% confidence interval for the estimated effect of the social distance manipulation did not exclude zero. Converting from log odds to *d*, the estimated effect for the social distance manipulation accounting for valence differences was $d = 0.006$, 95% CI [-0.05, 0.07].

Likelihood

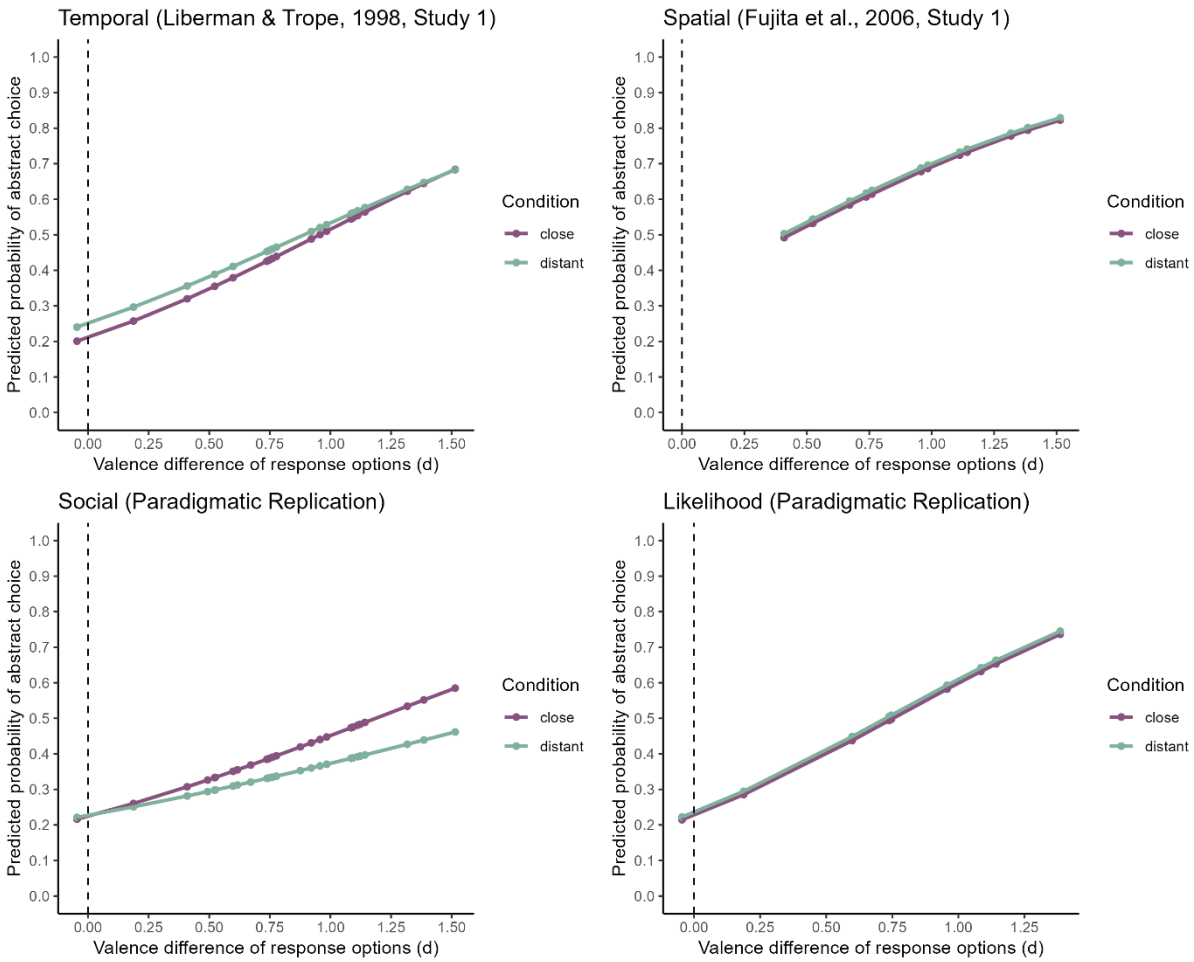
For the likelihood replications, likelihood ratio tests indicated that adding valence differences to the model offered significant improvement to the model, $\chi^2(1) = 14.26$, $p < .001$, and adding the interaction between the manipulation and valence differences did not offer significant improvement to the model, $\chi^2(1) = 3.00$, $p = .083$. In the retained model, the coefficient for the manipulation was $b = 0.05$, 95% CI [-0.04, 0.13]. The coefficient for valence differences was $b = 1.63$, 95% CI [1.08, 2.17].

To summarize, participants tended to select the more abstract option more frequently when it was more positive. At the average level of difference in valence between the response options ($d = 0.75$), the lower bound of the 95% confidence interval for the estimated effect of the likelihood manipulation did not exclude zero. Converting from log odds to *d*, the estimated

effect for the likelihood manipulation accounting for valence differences was $d = 0.03$, 95% CI [-0.02, 0.07].

Figure 5

Predicted Probability of Selecting the Abstract Option as a Function of Valence Differences, by Experiment



Note. Predicted probabilities at observed valence difference values (marked with points) are connected with interpolation lines.

Exploratory and Post Hoc Analyses

The following analyses were not pre-registered. Unless otherwise specified, exploratory analyses used the full sample, including data from participants who failed the comprehension check.

Country and Language Differences. The contributing labs originated from 27 countries and regions and used 15 languages, so it is worthwhile to investigate whether the effect sizes varied across these factors. We conducted exploratory analyses using country and language as random effects. However, we found no evidence that country or language influenced the effect size of any of the studies. In the interest of space, these results are presented in the supplemental material (https://osf.io/8djym). Given that there was no evidence of significant heterogeneity for any of the effects, this lack of influence by country and language is unsurprising.

Differences Across Modality (In-Lab vs. Online). It is possible that participating in a physical laboratory setting produces different effect sizes compared to participating online. We compared the effect sizes from in-lab and online data collections with a moderation analysis. Effect sizes from two labs were excluded because they switched from in-lab to online during data collection. There was no evidence that the effects significantly differed across modalities for any of the four experiments. These results are presented in more detail in the supplemental material (https://osf.io/7rd8b).

Robustness Check: Location Check for Spatial Distance Replication. The spatial distance study materials assumed that participants were in a specific city, but online data collections do not stop participation from other locations. In the materials for online data collections, we included a question asking if the participant was in the correct city. Of the 903 participants for whom we had data for this question, 101 (11.2%) reported being in the incorrect location. To assess whether the results differ when excluding data from people reporting being

in the incorrect location, we repeated the main analysis, the comprehension check robustness check, and the modality moderation with these participants removed. Three labs that collected data at least partially online were missing the location check question due to technical or procedural errors. The effect sizes of these labs were excluded from this analysis. These additional analyses produced results that were nearly identical to the main results. These results are reported in detail in the supplemental material (https://osf.io/bc8pd).

Cause Size and Effect Size. There was evidence of significant heterogeneity in the manipulation checks across labs for all four studies (see https://osf.io/x2bh7). If the strength of the manipulation (i.e., the "cause size"; Abelson, 1995; Ejelöv & Luke, 2020) varies, it is plausible that the effect size would positively correlate with manipulation strength, such that the effect is only present or is stronger when the manipulation is stronger. To test this possibility, we fit meta-regression models predicting the effect sizes from the cause sizes, and we found no evidence of a significant relationship between cause size and effect size. The results of these analyses are presented in the supplemental material (https://osf.io/ugvtp).

Item-Level Effects. It is possible that effect sizes vary across the different items of the BIF. To investigate this possibility, for each study, we calculated effect sizes for each BIF item for each lab and synthesized them in a mixed-effects meta-analytic model, accounting for each lab as a random effect and treating the items as a categorical moderator. The effect sizes for each item estimated by these models are presented in Figure 6. These results are presented in detail in the supplemental materials (https://osf.io/tuc7e). As can be seen, all the confidence intervals for the item-level effects for the spatial distance and likelihood studies included zero. In the social distance study, the upper bound of 15 out of 25 items' confidence intervals excluded zero, consistent with the overall (negative) effect, and one item's confidence interval excluded zero in the positive direction. Note that these effect sizes do not account for the valence differences in the response options. In the temporal distance study, five out of 19 items'

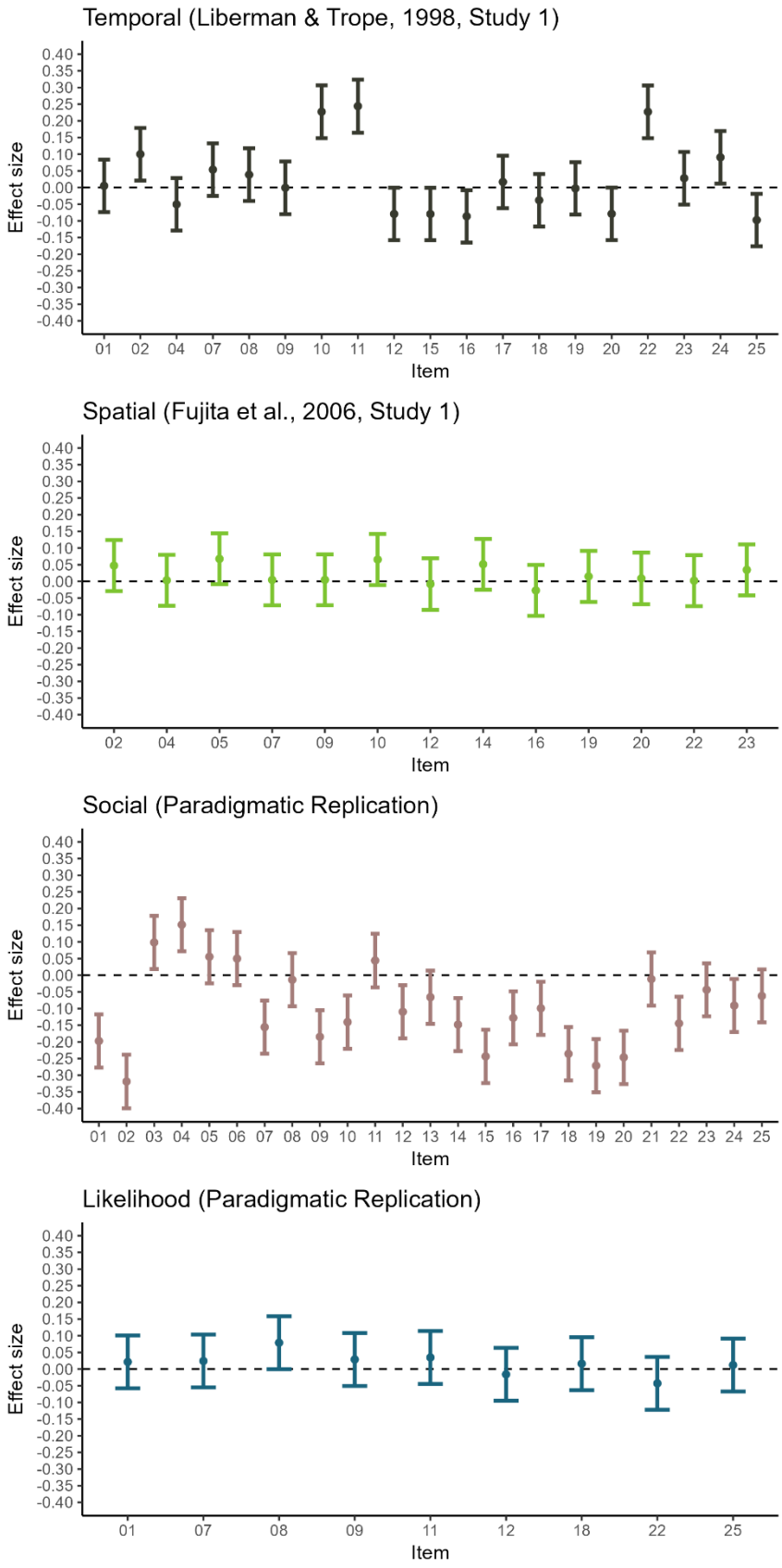
confidence intervals excluded zero in a direction consistent with the overall effect, and five items' confidence intervals excluded zero in the opposite direction.

Additional Exploratory Analyses

We conducted several exploratory analyses in addition to those described above. These include (but are not limited to) an examination of the potential moderating effects of positive and negative affect (measured by the PANAS), scores on the Analysis-Holism Scale, the physical distance between the cities used in the materials for the spatial distance experiment, the passage of time across the data collection period, and the amount of time taken by participants to complete the study. These analyses are documented in the supplemental material (https://osf.io/ra3dp/files/e5dc17d8-660e-4c68-a626-24eb38b195e0).

Figure 6

Item-Level Effects on the BIF



Note. Error bars represent 95% confidence intervals.

Discussion

The current multi-lab study tested the central tenet of construal level theory: that psychologically distant events are mentally construed more abstractly than psychologically near events. We tested this hypothesis by varying temporal, spatial, and social distance and the likelihood of events. Temporal distance and spatial distance were examined by direct replications of previously published studies, while social distance and likelihood were examined by paradigmatic replications. The present studies were selected and designed largely for their relevance to the fundamental hypotheses of CLT (e.g., using direct manipulations of psychological distance) so that their results would be theoretically informative (see e.g., Nosek & Errington, 2020).

Overall, results showed limited support for the predictions. According to our pre-registered criteria, the replication effect for temporal distance was consistent in direction, but inconsistent in magnitude, with the original effect reported by Liberman and Trope (1998, Study 1): The main analysis revealed an effect in the predicted direction with a confidence interval that excluded zero, but the observed effect (d = 0.08, 95% CI [0.003, 0.16]) was only 9% the size of the original effect (d = 0.92), and the upper bound of the replication confidence interval excluded the original effect size. Put differently, when participants were asked to imagine the activities on the BIF occurring “next year” rather than “tomorrow”, on average, they selected 0.32 more abstract options on the 19-item BIF. The replication effects for spatial distance and likelihood were small and had confidence intervals that included zero, failing to support the original predictions. Finally, the replication effect for social distance had a confidence interval that excluded zero, but the effect was in the opposite direction of the predicted effect, thus failing to support the prediction following from CLT. This effect was

eliminated by controlling for a confound in the response options of the BIF (see Limitations). The main analyses further revealed that none of the replication effects were associated with heterogeneity greater than what would be expected by random sampling error, indicating that the small meta-analytic effects cannot be attributed to the presence of potent lab-level moderators.

Our planned power analyses showed that the studies in the previous literature overall had virtually no statistical power to detect the effects implied by the replication studies, except for the social distance effect size (which was in the opposite direction to the prediction). Additionally, the median effect size for which previous CLT experiments had 80% power to detect was 8.5 times larger than the effect we observed for temporal distance (d = 0.08 vs. d = 0.68). To the extent that these replications are representative of studies in the CLT literature, the present results raise the concern that prior studies have been severely underpowered to detect and estimate relevant effects. This concern poses a threat to the overall validity of the previous literature, given that a literature based on small-sample studies is especially vulnerable to biasing influences, such as questionable research practices and publication bias (Bakker et al., 2012).

Limitations

The current replication studies relied solely on the Behavior Identification Form (BIF; Vallacher & Wegner, 1989) as the outcome measure. The interpretation of the replication results is thus contingent on the suitability of the BIF as a measure of mental abstraction. In the preparation of the current replication studies, the lead authors conducted extensive pretesting showing that the BIF indeed is highly sensitive to direct manipulations of mental abstraction (d = 1.42) and is superior to other considered outcome measures in this regard (Mac Giolla et al., 2024). These findings, in conjunction with the fact that it is the most frequently reported

outcome measure in the literature on mental abstraction (Burgoon et al., 2013), speak to the utility of the BIF for the purpose of the current replication studies. With that said, the BIF only captures some aspects of abstraction as conceptualized in CLT. In addition to the distinction between why and how an action is to be performed as operationalized in the BIF, CLT also describes abstract construals (as opposed to concrete construals) in terms of causes (as opposed to effects), ends focused (as opposed to means focused), and wide categories (as opposed to narrow exemplars; Liberman & Trope, 2014). The BIFs under coverage of abstraction constrains the generalizability of the current findings.

It should also be noted that a confound is built into several of the BIF items, such that people perceive the abstract (vs. concrete) action descriptions more positively (https://osf.io/g6d5v/). Our robustness analyses showed that although the effects for temporal distance, spatial distance, and likelihood remained relatively stable, the observed effect for social distance (opposite to the predicted direction) was eliminated entirely when controlling for response option valence. Thus, this effect was completely accounted for by the fact that, when presented with a socially close (vs. distant) target, participants tended to identify that person's actions using more positive descriptions. This finding highlights the need to develop valence-neutral measures of mental abstraction. Of further note, our exploratory analyses revealed that the replication effect for temporal distance was present in only five of the 19 BIF items included in the study, which calls for further examination of the generality of the effect. Newer measures may address some of the concerns with the BIF. For instance, recent research suggests that using a modified version of the BIF as the dependent variable may produce larger and more reliable effects of temporal distance on abstraction (Nguyen et al. 2023).

A second potential limitation is that approximately 17% of our participants failed the comprehension check. This rate varied little between online and in-lab administrations of the studies. As CLT research has not typically reported comprehension checks, we cannot know

whether a 17% failure rate is representative of the field. Mitigating this issue, results were largely unchanged when excluding participants who failed the check. One exception is that the confidence interval for the temporal distance effect size no longer excluded zero when excluding those who failed the comprehension check. In absolute terms, however, the point estimate and confidence bounds changed only a small amount.

Constraints on Generality

The contributing labs represent a diverse set of geographical locations, languages, and cultures. Thus, we do not consider a lack of diversity to pose a serious threat to the generality of our conclusions, although representation from Africa and South America is regrettably lacking. Additionally, the low amount of heterogeneity associated with the replication effects, despite the large variability in many lab-specific parameters, increases our confidence of the generalizability of the reported findings. That being said, the sample of participants consisted of mostly women—a limitation shared with much of CLT research (Soderberg et al., 2015).

Conclusion

The current research presents strong evidence of weak or non-existent relationships between the four forms of psychological distance and mental abstraction, as operationalized in the current studies. CLT holds that the relationship between psychological distance and mental abstraction constitutes a fundamental and universal mechanism of human cognition (see e.g., Liberman & Trope, 2014; Trope & Liberman, 2010). Given sensitive measures, strong manipulations, and high statistical precision, one would expect a fundamental process of human cognition, previously observed in sample sizes a fraction as large as the current ones, to produce effects greater than $d = 0.08$—the largest observed effect in the present studies. Because the present studies use methods representative of the literature, these findings—discrepant with the

broad predictions of the theory—present a challenge for CLT. A theory of CLT's breadth and influence should be able to better specify the conditions under which the hypothesized relationships can be reliably demonstrated through rigorous replications.

The current findings also raise a question with important applied implications: How can the direct effects of psychological distance on mental abstraction, estimated here to be very small at best, account for the large downstream consequences that have been documented in the existing literature (Soderberg et al., 2015)? We encourage researchers to consider whether alternative theoretical or methodological explanations, which do not require CLT as an explanatory framework, can provide plausible accounts of such findings. Finally, systematic validation is necessary to rule out the possibility that the current findings are simply due to a lack of adequate methods for manipulating and measuring the constructs of interest. We anticipate and hope that the current research will inspire theory development, methodological refinement, and a renewed focus on the practical applicability of CLT.

References

Abelson, R. P. (1995). *Statistics as principled argument*. Lawrence Erlbaum.

Adler, S., & Sarstedt, M. (2021). Mapping the jungle: A bibliometric analysis of research into construal level theory. *Psychology & Marketing*, *38*, 1367-1383. https://doi.org/10.1002/mar.21537

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Bartoš, F., Maier, M., Wagenmakers, E.J., Doucouliagos, H., & Stanley, T. D. (2021). No need to choose: Robust Bayesian meta-analysis with competing publication bias adjustment methods. *Research Synthesis Methods*. https://doi.org/10.31234/osf.io/kvsp7

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185-216. https://doi.org/10.1177/135910457000100301

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*(5), 501-520. https://doi.org/10.1177/1745691613497964

Calderon, S., Mac Giolla, E., Ask, K., & Granhag, P. A. (2020). Subjective likelihood and the construal level of future events: A replication study of Wakslak, Trope, Liberman, and Alony (2006). *Journal of Personality and Social Psychology*, *119,* e27-e37. https://doi.org/10.1037/pspa0000214

Danziger, S., Monit, R., & Barkan, R. (2012). Idealistic advice vs. pragmatic choice: A psychological distance account. *Journal of Personality and Social Psychology*, *102*, 1105-1117. https://doi.org/10.1037/a0027013

Ejelöv, E., & Luke, T. J. (2020). "Rarely safe to assume": Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937. https://doi.org/10.1016/j.jesp.2019.103937

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

Eyal, T., Sagristano, M. D., Trope, Y., Liberman, N., & Chaiken, S. (2009). When values matter: Expressing values in behavioral intentions for the near vs. distant future. *Journal of Experimental Social Psychology*, *45*(1), 35-43. https://doi.org/10.1016/j.jesp.2008.07.023

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*(3), 313-332. https://doi.org/10.1016/j.jesp.2008.07.023

Fujita, K., Henderson, M. D., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science*, *17(4)*, 278-282. https://doi.org/10.1111/j.1467-9280.2006.01698.x

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Gong, H., & Medin, D. L. (2012). Construal levels and moral judgment: Some complications. *Judgment and Decision Making, 7*(5), 628-638. https://doi.org/10.1017/S1930297500006343

Grinfeld, G., Wakslak, C. J., Trope, Y., & Liberman, N. (2021, April 7). *Hypotheticality and level of construal*. https://doi.org/10.31234/osf.io/yvafk

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas,

C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., … Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*(1), 159–169. https://doi.org/10.1038/s41562-020-01007-2

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences*, *18*(7), 364–369. https://doi.org/10.1016/j.tics.2014.03.001

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, *75*, 5-18. https://doi.org/10.1037/0022-3514.75.1.5

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*(5), 1256-1269. https://doi.org/10.1016/j.jesp.2008.04.007

Luke, T. J., Ask, K., Magnusson, E., Calderon, S., & Mac Giolla, E. (2021). *Revisiting the relationship between social distance and communication preferences: A replication and reinterpretation of Amit et al. (2013, Experiment 2).* https://doi.org/10.31234/osf.io/qjw7g

Mac Giolla, E., Luke, T. J., Calderon, S., & Ask, K. (2024). *Validating measures of mental abstraction*. https://doi.org/10.31234/osf.io/v6xt4

Maier, M., Bartoš, F., Oh, M., Wagenmakers, E. J., Shanks, D., & Harris, A. J. L. (2024). Publication bias in research on construal level theory. https://doi.org/10.31234/osf.io/r8nyu

Martín-Fernández, M., Requero, B., Zhou, X., Gonçalves, D., & Santos, D. (2022). Refinement of the Analysis-Holism Scale: A cross-cultural adaptation and validation of two shortened measures of analytic versus holistic thinking in Spain and the United

States. *Personality and Individual Differences*, *186*, 111322. https://doi.org/10.1016/j.paid.2021.111322

McCarthy, R. J., Hartnett, J. L., Heider, J. D., Scherer, C. R., Wood, S. E., Nichols, A. L., ... & Walker, W. R. (2018). An investigation of abstract construal on impression formation: A multi-lab replication of McCarthy and Skowronski (2011). *International Review of Social Psychology, 31*(1). https://doi.org/10.5334/irsp.133

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353-383. https://doi.org/10.1016/0010-0285(77)90012-3

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Nussbaum, S., Liberman, N., & Trope, Y. (2006). Predicting the near and distant future. *Journal of Experimental Psychology: General, 135*(2), 152-161. https://doi.org/10.1037/0096-3445.135.2.152

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sánchez, A. M., Coleman, C. W., & Ledgerwood, A. (2021). Does temporal distance influence abstraction? A large pre-registered experiment. *Social Cognition*, *39*(3), 352-365. https://doi.org/10.1521/soco.2021.39.3.352

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80. https://doi.org/10.1177/1745691613514755

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*(3), 525-548. https://doi.org/10.1037/bul0000005

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403-421. https://doi.org/10.1037/0033-295X.110.3.403

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440-463. https://doi.org/10.1037/a0018963

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660-671.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

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*(4), 641-653. https://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*(6), 1063-1070.

Westfall, J. (2016). PANGEA: Power analysis for general ANOVA designs. *Unpublished manuscript.* Available at http://jakewestfall.org/publications/pangea.pdf

Yan, D., Sengupta, J., & Hong, J. (2016). Why does psychological distance influence construal level? The role of processing mode. *Journal of Consumer Research*, *43*, 598-613. https://doi.org/10.1093/jcr/ucw045

Žeželj, I. L., & Jokić, B. R. (2014). Replication of experiments evaluating impact of psychological distance on moral judgment. *Social Psychology*, *45*, 223-231. https://doi.org/10.1027/1864-9335/a000188

Appendix A: Items of the Behavior Identification Form

Table A1

Items of the Behavior Identification Form (BIF) Included in the Primary Analysis for Each Study

	Included in the primary analysis			
Item	Temporal	Spatial	Social	Likelihood
1. Making a list	Yes	-	Yes	Yes
2. Reading	Yes	Yes	Yes	-
3. Joining the army	-	-	Yes	-
4. Washing clothes	Yes	Yes	Yes	-
5. Picking an apple	-	Yes	Yes	-
6. Chopping down a tree	-	-	Yes	-
7. Measuring a room for carpeting	Yes	Yes	Yes	Yes
8. Cleaning the house	Yes	-	Yes	Yes
9. Painting a room	Yes	Yes	Yes	Yes
10. Paying the rent	Yes	Yes	Yes	-
11. Caring for houseplants	Yes	-	Yes	Yes
12. Locking a door	Yes	Yes	Yes	Yes
13. Voting	-	-	Yes	-
14. Climbing a tree	-	Yes	Yes	-
15. Filling out a personality test	Yes	-	Yes	-
16. Brushing teeth	Yes	Yes	Yes	-
17. Taking a test	Yes	-	Yes	-
18. Greeting someone	Yes	-	Yes	Yes
19. Resisting temptation	Yes	Yes	Yes	-
20. Eating	Yes	Yes	Yes	-
21. Growing a garden	-	-	Yes	-
22. Traveling by car	Yes	Yes	Yes	Yes
23. Having a cavity filled	Yes	Yes	Yes	-
24. Talking to a child	Yes	-	Yes	-
25. Pushing a doorbell	Yes	-	Yes	Yes

Appendix B: Comprehension Checks

Response options for all comprehension checks were presented to participants in random order.

Temporal Distance

You were previously asked to imagine engaging in a series of activities (e.g., making a list, painting a room, brushing teeth). **When were these events supposed to take place?**

1. Next year [correct answer distant condition]
2. Tomorrow [correct answer close condition]
3. In five weeks
4. In six months
5. 18 months from now
6. In two years

Spatial Distance

You were previously asked to imagine that you had moved to a new place. **Where was this new place located?**

1. Outside of [current location], about [3 miles/5 km] from here [correct answer close condition]
2. Outside of [distant location], about [x] [miles/km] from here [correct answer distant condition]
3. Outside of [current location], about [15 miles/25 km] from here
4. Outside of [alternative distant location], about [x] [miles/km] from here
5. In [a different country closest to the current location]
6. In [a different country far away from the current location]

Social Distance

You were previously asked to imagine a person engaging in a series of activities (e.g., making a list, painting a room, brushing teeth). **What was the person's name?**

1. [Target Name] [correct answer]
2. [Filler Name 1]
3. [Filler Name 2]
4. [Filler Name 3]
5. [Filler Name 4]
6. [Filler Name 5]

Likelihood

You were previously asked to imagine helping your friend move. **In the scenario, how likely was it that your friend was going to move?**

1. There was a 5% chance [correct answer in the distant condition]
2. There was a 10% chance
3. There was a 25% chance
4. There was a 50% chance
5. There was a 75% chance
6. There was a 95% chance [correct answer in the close condition]

Appendix C: Manipulation checks

Temporal Distance

As a final question, we are interested in how people perceive temporal distances. Events can feel closer or more distant in time from the present moment. When events feel like they will happen soon, they are said to be temporally close. In contrast, when events feel like they will not happen for a long time, they are said to be temporally distant.

To what extent does something taking place [tomorrow]/[next year] feel temporally close or distant? (1 = *very close*, 7 = *very distant*)

Spatial Distance

As a final task, please take a moment and imagine that you have moved to a new place just outside of [current location, just under 3 miles/5 km]/[distant location, over X miles/km] away.

Places can feel close to or far away from ourselves in a psychological sense. Places that feel like they are close to where you are, are said to be "spatially close". In contrast, places that feel like they are very far away, are said to be "spatially distant".

To what extent does a place located outside of [current location, just under 3 miles/5 km]/[distant location, over X miles/km] away, feel spatially close or distant? (1 = *very close*, 7 = *very distant*)

Social Distance

As a final question, we are interested in how people form impressions of others based on minimal information. Below is a brief description of a person. Please read the description and try to imagine the described person. We realize that it is not possible to form an accurate impression based on so little information. But here we are interested in your own subjective perception and the type of person that you spontaneously imagine when reading the description. In other words, there are no correct or incorrect answers.

[Name] is [a man]/[a woman]/[][4] aged [x]. [Name] has an educational background that is [similar to]/[very different from] yours and [he]/[she]/[][5] [shares several of]/[does not share any of] your personal interests. In other words, [Name] is a person with whom you have [a lot]/[little] in common.

People can feel close to or far away from ourselves in a psychological sense. When a person feels psychologically close to ourselves, they are said to be "socially close". In contrast, when a person feels psychologically removed from ourselves, they are said to be "socially distant".

[4]In the socially close condition, no gender was specified for participants identifying with non-binary or other genders.

[5]In the socially close condition, no pronoun was specified for participants identifying with non-binary or other genders.

To what extent do you feel socially close or socially distant to [Name]? (1 = *very close*, 7 = *very distant*)

Likelihood

As a final task, please take a moment and imagine that a friend has asked if you can help them move into a new apartment. The friend will only be moving if he is offered a new job he has applied for. The friend thinks there's a [95%]/[5%] chance that he will be offered the job. In other words, you probably [will]/[won't] have to help him move.

Events can feel more or less close or distant in a psychological sense because they feel more or less likely to happen. An event may feel close because it feels likely to happen. In contrast, an event may feel distant because it feels unlikely to happen.

To what extent does a [95%]/[5%] chance of helping your friend move feel close or distant in terms of likelihood? (1 = *very close*, 7 = *very distant*)