Replication of Experiments Evaluating Impact of Psychological Distance on Moral Judgment
(Eyal, Liberman & Trope, 2008; Gong & Medin, 2012)

Iris L. Žeželj, and Biljana R. Jokić

Department of Psychology, University of Belgrade, Serbia

Abstract. Eyal, Liberman, and Trope (2008) established that people judged moral transgressions more harshly and virtuous acts more positively when the acts were psychologically distant than close. In a series of conceptual and direct replications, Gong and Medin (2012) came to the opposite conclusion. Attempting to resolve these inconsistencies, we conducted four high-powered replication studies in which we varied temporal distance (Studies 1 and 3), social distance (Study 2) or construal level (Study 4), and registered their impact on moral judgment. We found no systematic effect of temporal distance, the effect of social distance consistent with Eyal et al., and the reversed effect of direct construal level manipulation, consistent with Gong and Medin. Possible explanations for the incompatible results are discussed.

Keywords: psychological distance, moral judgment, construal level theory, replication

Most of the time people act and feel toward the construal of an object, rather than object itself. Construal Level Theory (CLT; Trope & Liberman, 2010) addresses universally relevant issues of mental representations and thus runs across the subdisciplines of psychology. CLT has been widely tested in different domains, and seems to have received relatively robust empirical confirmation. It is established as an influential theory of social cognition: A chapter is regularly devoted to it in contemporary handbooks (Shapira, Liberman, Trope, & Rim, 2012; Trope & Liberman, 2011), and the seminal article on Temporal construal (Trope & Liberman, 2003) has been highly cited (656 citations in Scopus, 689 in APA Psych Net, and 1,324 times in Google scholar[1]).

CLT's major premise is that mental representations of objects and situations vary depending on their distance from the perceiver: Distant objects tend to be represented by few essential characteristics (high-level construal), whereas close objects tend to be represented by detailed and contextual information (low-level construal). High-level construals are important for regulation toward distal objects, when it is important to focus on their invariant, central features, while low-level construals are more important for regulation toward close objects, as we need to act toward them immediately. Given that psychological distance is egocentric, with its reference point being here and now, it can be social (self vs. other), temporal (present

self vs. future or past self), spatial (here vs. some other place), or hypothetical (highly likely vs. unlikely event).

CLT proposes that moral values have stronger impact on judgment from a greater psychological distance, while contextual information is more relevant when at a close distance. In other words, the same act will be evaluated differently depending on the perceivers' distance from the presented event. In a series of experiments, Eyal, Liberman, and Trope (EL&T, 2008) tested how temporal and social distance impact the evaluation of moral acts (both transgressions and virtuous acts), followed by contextual information that intended to attenuate their severity. Temporal distance was manipulated by the instruction to imagine the situation was taking place now or later in time. Social distance was manipulated by the instruction to imagine the situation from one's own or from a third person's perspective. Four experiments in their article confirmed CLT propositions: Both moral transgressions and virtuous behaviors were evaluated differently depending on the psychological distance. When the distance was higher, values (as high-level construal) had a greater impact on evaluation. When the distance was smaller, contextual information (as low-level construal) was of greater concern. This study has been cited (33 times in Scopus, 76 in Google scholar), as supportive of CLT predictions within the moral evaluation domain. In a similar vein, Agerström and Bjorklund (2009) examined the impact of temporal distance on

---

moral concerns (selfish vs. altruistic considerations) and reported that moral concerns were higher for temporarily distant situations. In a series of experiments, Lammers' results (2012) mirrored those of EL&T: Subjects reacted more negatively to others' morally questionable behaviors when adopting an abstract rather than concrete perspective.

A following project by Gong and Medin (G&M, 2012), however, yielded some directly conflicting results. The authors used the same stimulus material as EL&T (vignettes about moral transgressions and virtuous acts), but instead of manipulating distance, they directly manipulated construal level by priming tasks designed to stimulate either the abstract or concrete mindset. Contrary to expectations, when participants were primed with a concrete mindset, they showed more extreme moral judgments of both moral transgressions and virtuous behaviors. In order to test the possible impact of a different procedure (manipulation of distance vs. manipulation of construal level), G&M directly replicated one of four experiments from EL&T's study. Again, G&M observed the opposite results. When EL&T tried to replicate G&M's experiments (using the priming tasks), their results were not significant. Baring in mind these contradictory findings and the importance of the topic (impact of construal level on moral judgment), we proposed a two-stage replication project:

1. Direct replication of EL&T's Studies 2, 3, and 4 and G&M's Study 1 in the Social Cognition Laboratory at Belgrade University, Serbia. The main authors of both studies agreed to provide us with the materials (vignettes, instruments, and instructions) used in their experiments.
2. Aggregation of databases from three laboratories: Ben-Gurion University (Israel), Northwestern University (Ilinois, USA), and Belgrade University (Serbia). This provided us with the opportunity to directly compare: (a) The effects obtained by three laboratories (Experiment 2 from EL&T's study, and Experiment 1 from G&M's study); (b) the effects obtained by two laboratories (Israeli and Serbian), from Experiments 3 and 4 of EL&T's.

Materials, data, and the preregistered proposal are available on the project page on the Open Science Framework (https://osf.io/Z5TE6/).

We report all data exclusions, manipulations, and measures, and how we determined our sample sizes. The planned sample size[2] was based on the effect size from each of the original studies, so as to achieve .95 power (see Table 1 for details).

Whenever we had an opportunity, we tested up to 5% respondents over the planned sample size, in case we had to omit some from further analysis. Participants were randomly assigned to experimental groups. They were recruited from a pool of psychology students from Belgrade

Table 1. Effect sizes from original studies and planned sample size

Original study	Original sample size	Cohen-s' d from original study	Planned sample size
Exp 2 (EL&T)	58	.68	114 (57 per group)
Exp 3 (EL&T)	40	.72	102 (51 per group)
Exp 4 (EL&T)	47	.81	80 (40 per group)
Exp 1 (G&M)	34	1.07	48 (24 per group)

University, in exchange for course credits. The same recruiting method was applied in EL&T's original study, while in G&M's study participation was on a voluntary basis. Gender information was registered, although it was not found to have a significant effect in either study.

We used the vignettes from EL&T's studies, with the authors' permission. The vignettes were translated into Serbian by two independent bilingual translators, and then translated back into English in order to provide maximum correspondence (as suggested in Brislin, 1970, 1976). As was the procedure in the original experiments, the respondents filled in booklets in paper format.

The experimenters were PhD students from the Social Cognition Laboratory of the Department of Psychology, Belgrade University, blind to the study hypothesis. Participants were tested in groups no larger than ten. At the beginning of each session, the experimenter presented themselves to the participants and explained the research purpose: "This study is about a judgment of different people's actions. After reading each story, please provide your opinion below it. There are no right or wrong answers. You are expected to evaluate just as you think. Examination is anonymous, so you do not need to provide any personal information." The experimenter was present until the end of the experiment, but not allowed to give any additional instructions, except to encourage participants to give their own opinions if they had any questions. After hearing the instructions, participants read the vignettes and evaluated the wrongness (in the first two studies), virtuousness (in the third study), or moral acceptability (in the fourth study) of the actions presented. We used the same scales as in corresponding studies of EL&T and G&M.

Our analysis was planned in a confirmatory fashion.[3] We performed a series of standard *t*-tests, as well as default Bayesian *t*-tests, as proposed by Rouder, Speckman, Sun, Morey, and Iverson (2009), in order to test the differences:

1. In perceived wrongness of the actions between high and low temporal distance primed groups (Study 1);
2. In perceived wrongness of the actions between high and low social distance primed groups (Study 2);
3. In perceived virtuousness of the actions between high and low temporal distance primed groups (Study 3);

---

4. In perceived moral acceptability of the transgressions between high and low construal level primed groups (Study 4).

In a next step we performed an analysis on the aggregated database (with provided data from two or three laboratories), and at the end we introduced "laboratory" as a factor.

The main known difference from the original study was the cultural and linguistic backgrounds of the samples (Serbian vs. Israeli/Hebrew vs. American/English). One minor difference was that our sample consisted exclusively of students, while in one of EL&T's Study 3 participants were workers from security service organizations. Apart from that, a full methodological and procedural equivalence was set up.

Study 1

The aim of the Study 1 was to replicate the findings of EL&T (Study 2, 2008) indicating that people would judge immoral acts more harshly if presented to them as temporally distant rather than presented as temporally close. Participants judged the wrongness of moral transgressions as expected to occur either the next day (near future condition) or next year (distant future condition).

Method

Participants

Participants in our study were 116 undergraduate students from University of Belgrade, Serbia, who participated in exchange for course credit. Our aggregated database included 58 participants from the original Israeli study and 36 from American replication; a total of 210 participants.

Procedure

Participants read three vignettes (adopted from Haidt 2001; Haidt, Koller, & Dias, 1993; as in EL&T, 2008) describing a moral transgression followed by situational details that moderated the offensiveness of the action, for example "sister and brother had a sexual intercourse" (violating a widely accepted moral rule), but "they used birth control and agreed to do it only once" (contextual information that was supposed to attenuate the severity of the act). They were asked to imagine that the events would happen the next day or the next year. After reading each vignette, participants evaluated the wrongness of the actions on a scale ranging from −5 (= *very wrong*) to +5 (= *completely acceptable*).

Results for Serbian Replication

A mixed ANOVA with temporal distance (near vs. distant future) as between-subject factor and story (eating one's dead pet, sexual intercourse with sibling, dusting with national flag) as within-subject factor yielded a main effect of story, $F(2, 115) = 167.29$, $p < .001$, $\eta^2 = .59$, indicating that the wrongness of events were judged differently ("incest": $M = -3.93^4$, $SD = 0.21$; "dog": $M = -3.82$, $SD = 0.19$; "flag": $M = 0.59$, $SD = 0.32$).

More importantly and in contrast to original EL&T's study (Study 2, EL&T, 2008), there was no main effect of temporal distance, $F(1, 115) = 0.11$, $p = .746$, $\eta^2 = .001$, $g = -.06$, [CI = −0.428, 0.310], meaning that distant future transgressions were judged just as unacceptable ($M = -2.32$; $SD = 2.13$) as near future transgressions ($M = -2.44$; $SD = 1.91$). Across three stories, there was a marginally significant effect of distance only on one ("dog," $M_{close} = -4.19$, $SD = 1.51$; $M_{dist} = -3.44$, $SD = 2.48$, $F(1, 115) = 3.87$, $p = .052$, $\eta^2 = .03$).

Scaled JZS Bayes Factor (1.12) supported the null hypothesis, indicating that two temporal distance groups did not differ from each other.[5]

Results for Aggregated Data From Three Laboratories

A mixed ANOVA with temporal distance (near vs. distant future) as between-subject factor and story (1–3) as within-subject factor again yielded a strong main effect of story, $F(2, 209) = 186.61$, $p < .001$, $\eta^2 = .47$ ("incest": $M = -4.06$, $SD = 0.14$; "dog": $M = -3.93$, $SD = 0.13$; "flag": $M = -0.50$, $SD = 0.24$), meaning that wrongness of different transgressions was judged differently. As for the central research hypothesis, we ended up with conflicting findings: The original study found the expected impact of temporal distance to the wrongness assessment; the direct replication with the American sample yielded the opposite pattern of results, whereas direct replication with the Serbian sample yielded no significant effect. Analysis on the integrated sample revealed no effect of temporal distance $F(1, 209) = 0.25$, $p = .618$ $\eta^2 = .001$, $g = -0.069$, [CI = −0.323, 0.186] (distant future transgressions: $M = -2.76$; $SD = 2.03$; near future transgressions: $M = -2.89$; $SD = 1.72$).

Further analysis revealed the significant effect of "laboratory" as a factor, $F(2, 209) = 11.80$, $p < .001$, $\eta^2 = .10$ and a significant interaction between story and

---

[4] We did not multiply the wrongness assessments by −1, as in in Eyal et al. study, since we obtained larger dispersion of raw scores, out of which some were positive and in the vignette "flag" the mean score was even positive.

[5] Bayes factors were calculated using online calculator provided by University of Missouri. It can be retrieved at http://pcl.missouri.edu/bayesfactor
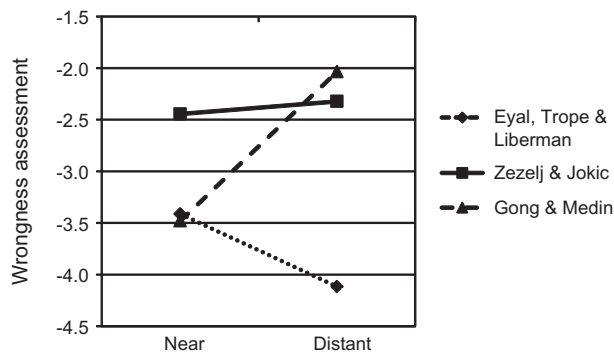
Figure 1. Acceptability of moral transgressions by temporal distance in three studies.

Figure 2. Wrongness of moral transgressions by social distance across stories.

laboratory, $F(4, 209) = 10.76$, $p < .001$, $\eta^2 = .09$, indicating cultural differences in assessment of wrongness of different immoral acts.[6] More importantly, there was also an interaction between temporal distance and "laboratory", $F(2, 209) = 4.08$, $p = .018$, $\eta^2 = .04$. A post hoc Tukey test showed that the Israeli laboratory differed significantly from both Serbian one (at $p < .001$) and American one (at $p = .022$), while Serbian and American laboratories did not differ from one another (see Figure 1).

Study 2

Study 2 was designed to build upon the results of Study 1 by manipulation of social distance (self vs. other) rather than temporal, and to extend the number of acts to be evaluated (six instead of three). It presents a direct replication of EL&T's study 3.

Method

Participants

Participants were 105 undergraduate students from University of Belgrade, Serbia. The original Israeli study included 40 participants, so aggregated database consisted of 145 participants.

Procedure

Participants read six vignettes. Out of those, two were the same as in Study 1 (eating a dog, dusting with a national flag) and the additional four were adopted from Haidt et al. (1993) (a girl pushing another kid off a swing, cousins kissing each other on the mouth, a man breaking a promise
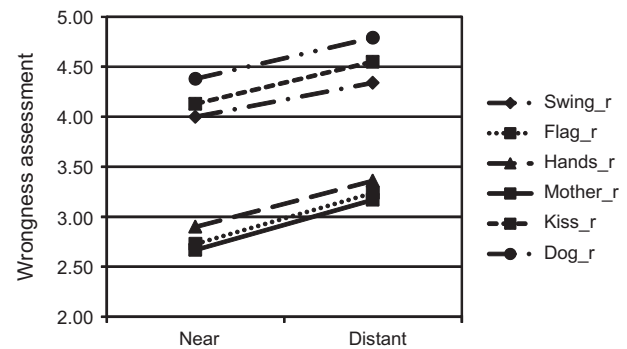
to his dying parent, and a man who ate with his hands in public). Participants were asked to think about a specific person they knew (high social distance condition) or to focus on their own feelings and thoughts (low social distance condition), and to judge the events either from a third person viewpoint or from a first person viewpoint. They evaluated the wrongness of the actions on a scale ranging from 1 (*not acceptable*) to 5 (*completely acceptable*).

Results for Serbian Replication

In order to unify data from replication and original databases, we reversed the participants' ratings so that higher ratings indicated higher wrongness scores. A mixed ANOVA with social distance (self vs. other) as between-subject factor and story (1–6) as within-subject factor yielded a main effect of story, $F(5, 104) = 52.01$, $p < .001$, $\eta^2 = .34$, indicating that the wrongness of events were judged differently. More importantly, there was also a main effect of social distance, $F(1, 104) = 9.90$, $p = .002$, $\eta^2 = .09$, $g = 0.615$, [CI = 0.479, 0.751]. All actions were judged as more wrong from a third person perspective ($M = 3.91$, $SD = 0.61$) then from a first person perspective ($M = 3.47$, $SD = 0.80$), which was in accordance with CLT predictions and the results of EL&T's original Study 3. Across six scenarios, the effect was significant on two ("dog", $F(1, 104) = 5.17$, $p = .025$, $\eta^2 = .05$; "broken promise", $F(1, 104) = 4.73$, $p = .032$, $\eta^2 = .04$), and marginally significant on three ("swing", $F(1, 104) = 3.39$, $p = .068$, $\eta^2 = .03$; "flag", $F(1, 104) = 3.34$, $p = .070$, $\eta^2 = .03$; "kiss", $F(1, 104) = 2.89$, $p = .092$, $\eta^2 = .03$). The pattern of means of six stories across social distance is presented in Figure 2.

Scaled JZS Bayes Factor which was lower than one (0.07) strongly supported the alternative hypothesis, indicating that two social distance groups significantly differed from each other.

---

[6]   This difference largely stems from different assessment in the vignette "flag": while Israeli and American participants thought this was an immoral act ($M = -2.48$, $SD = 2.35$; $M = -0.83$, $SD = 3.45$), Serbian participants even viewed it as somewhat positive ($M = 0.59$, $SD = 3.49$). As we anticipated this might happen, we made a note explaining potential reasons in our preregistration proposal (https://osf.io/qhagec).
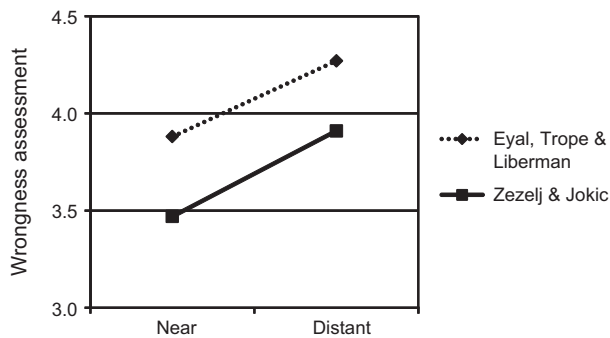
Figure 3. Wrongness of moral transgressions by social distance in two studies.

Results for Aggregated Data From Two Laboratories

As expected, mixed ANOVA with social distance (self vs. other) as between-subject factor and story (1–6) as within-subject factor again yielded a main effect of story, $F(5, 144) = 68.83$, $p < .001$, $\eta^2 = .33$. The main effect of social distance was once again significant, $F(1, 144) = 13.77$, $p < .001$, $\eta^2 = .09$, $g = 0.624$ [CI = 0.513, 0.736] (means are detailed in Figure 3).

Further analysis revealed the main effect of laboratory as a factor, $F(1, 144) = 9.36$, $p = .003$, $\eta^2 = .06$. There was no interaction between laboratory and social distance factors, and as in previous study, there was a significant interaction between story and laboratory that emerged, $F(5, 144) = 9.05$, $p < .001$, $\eta^2 = .06$.

Study 3

The main objective of Study 3 was to explore the effects of temporal distance on moral judgment of virtuous acts performed under attenuating circumstances (e.g., a fashion company donates to the poor and it positively affects its sales rate). In EL&T's original Study 4, higher distance lead to more positive virtuousness ratings, that is ascribing less weight to attenuating contextual information.

Method

Participants

Participants were 84 undergraduate students from Belgrade University, Serbia. They were randomized into ''near'' or ''distant'' future condition. As the original Israeli study included 47 participants, the aggregated database comprised of 131 in total.

Procedure

Participants were presented with three vignettes describing virtuous acts followed by extenuating contextual informa-

tion. They were asked to imagine a described event occurring the next day (low temporal distance) or in a year (high temporal distance). After that they evaluated the virtuousness of each act on a scale anchored with 1 (*not at all virtuous*) to 7 (*extremely virtuous*).

Results for Serbian Replication

We conducted a mixed ANOVA with virtuousness ratings as a dependent variable, temporal distance (near vs. distant future) as a between-subject factor and story (1–3) as a within-subject factor. Results yielded a main effect of story, $F(2, 83) = 16.68$, $p < .001$, $\eta^2 = .17$, indicating that the virtuousness of events was judged differently. Same as in our Study 1 and in contrast to EL&T's original Study 4, there was no main effect of temporal distance, $F(1, 83) = 1.46$, $p = .23$, $\eta^2 = .02$, $g = -0.261$, [CI = −0.521, −0.002], meaning that the virtuousness of distant future acts ($M = 4.55$; $SD = 1.27$) was judged the same as the virtuousness of near future acts ($M = 4.87$; $SD = 1.15$), with no significant differences across the stories.

Scaled JZS Bayes Factor higher than one (3.04) supported the null hypothesis, indicating that two temporal distance groups did not differ from one another.

Results for Aggregated Data From Two Laboratories

A mixed ANOVA with temporal distance (near vs. distant future) as between-subject factor, and story (1–3) as within-subject factor again yielded a main effect of story, $F(2, 130) = 14.97$, $p < .001$, $\eta^2 = .10$. There was no main effect of temporal distance, $F(1, 130) = 0.11$, $p = .737$, $\eta^2 = .001$, $g = 0.061$, [CI = −0.136, 0.257] meaning that the virtuousness of distant future acts ($M = 4.61$; $SD = 1.13$) was judged the same as the virtuousness of near future acts ($M = 4.54$; $SD = 1.17$).

Further analysis revealed the marginally significant effect of laboratory as a factor, $F(1, 130) = 3.31$, $p = .071$, $\eta^2 = .025$ (see Figure 4). There was also an interaction between temporal distance and laboratory, $F(1, 130) = 6.79$, $p = .01$, $\eta^2 = .05$, and as in previous studies, an interaction between story and laboratory, $F(2, 130) = 4.80$, $p = .009$, $\eta^2 = .04$.

Study 4

In this study we introduced a direct manipulation of construal level as employed in G&M's (2012) first study. We primed our participants with a series of ''how-and-why'' questions that were expected to directly activate either a low or high construal mindset. By asking subjects to generate subordinate goals, we led them to adopt an instrumental, lower-level perspective. In contrast, generating superordinate goals led them to adopt a higher-level perspective. We aimed to explore if priming a high-or-low construal
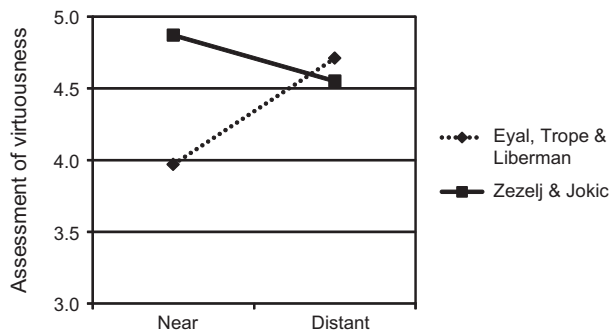
Figure 4. Virtuousness of moral acts by temporal distance in two studies.

level would have an impact on the severity of our participants' moral judgment. In G&M's study, this manipulation yielded results contradictory to those of EL&T: Low-level construals led to harsher condemnation of moral transgressions in comparison to high-level construals.

Method

Participants

Participants were 48 undergraduate students from the University of Belgrade, Serbia, randomly assigned to either how or why priming condition. There were 34 participants in G&M's original study, while in EL&T's replication study there were 81, which left our aggregated database with 163 participants.

Procedure

Participants completed both priming and evaluation task in one session, ostensibly as two independent studies. Half of the participants was asked *how* they could improve and maintain health, while the other half was asked *why* they should improve and maintain health. After stating the first reason, they were asked to respond to that reason in the same vain (i.e., how or why). They repeated this process four times, filling in a diagram. Upon finishing this task, participants were presented with four scenarios of moral violation (three vignetess were the same as in Study 1 of this paper, plus one additional of a student cheating in an exam). Respondents rated moral acceptability of each act on an 11-point scale ranging from −5 (= *extremely unacceptable*) to 5 (= *extremely acceptable*).

Results for Serbian Sample

We conducted a two (high vs. low level construal) by four (story 1–4) mixed design ANOVA on moral acceptability ratings, with construal as between-subjects and story as within-subjects factor. Results yielded a main effect of
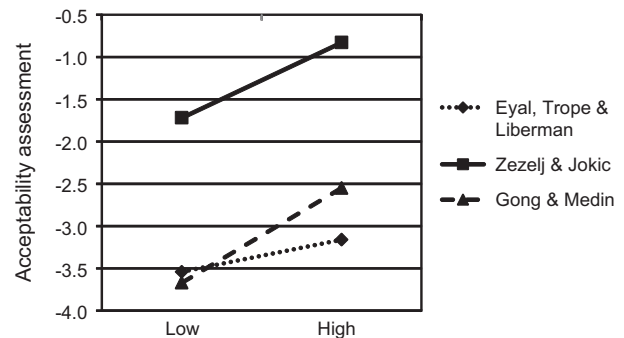

Figure 5. Acceptability of moral transgressions by construal level in three studies.

story, $F(3, 47) = 85.31$, $p < .001$, $\eta^2 = .65$. More importantly, there was also a main effect of construal level, $F(1, 47) = 5.68$, $p = .021$, $\eta^2 = .11$, $g = -0.681$, [CI = −1.045, −0.317], meaning that the wrongness of transgressions was judged in accordance to G&M's results. Participants judged the acts more harshly after low-level construal priming ($M = -1.72$, $SD = 1.55$), then after high-level construal priming ($M = -0.83$, $SD = 0.95$). Across four stories, the effect was significant in one ("flag," $M_{low} = 1.25$, $SD = 3.54$; $M_{high} = 3.50$, $SD = 2.00$, $F(1, 47) = 7.34$, $p = .009$).

Scaled JZS Bayes Factor (0.37) indicated scarce evidence for the alternative hypothesis.

Results for Aggregated Data From Three Laboratories

A mixed ANOVA with construal level (low vs. high) as between-subject factor and story (1–4) as within-subject factor again yielded a main effect of story, $F(3, 162) = 54.56$, $p < .001$, $\eta^2 = .25$. There was also a main effect of the construal level, $F(1, 162) = 8.81$, $p = .003$, $\eta^2 = .05$, $g = -0.338$, [CI = −0.564, −0.113] in contrast to CLT prediction: Low level priming led to harsher wrongness assessment ($M = -3.03$, $SD = 1.41$) than high level priming ($M = -2.35$, $SD = 1.53$).

Further analysis revealed the significant effect of the laboratory, $F(2, 162) = 51.78$, $p < .001$, $\eta^2 = .40$. A post hoc Tukey test demonstrated that the Serbian laboratory differed significantly from both Israeli and American, at $p < .001$), whereas Israeli and American laboratories did not differ from one another (as can be seen in Figure 5). Once more we discovered an interaction between story and laboratory, $F(6, 162) = 26.98$, $p < .001$, $\eta^2 = .26$. The difference was mainly due to the fact that Serbian participants judged two transgressions less harshly than the Israeli and American: "flag" and "cheat" (both at $p < .001$). We have already addressed the former; the latter might be because the concept of academic honesty is more vague in Serbian than in Israely/American university setting, with no explicit (written) ethical guidelines.

Table 2. Results summary of multiple studies

Study	Eyal, Liberman, and Trope, Hedge's g ± CI[7]	Zezelj and Jokic, Hedge's g ± CI	Gong and Medin, Hedge's g ± CI	Do the studies agree about the direction of the effect?	What is the pattern of statistical significance?	Is the effect size from original study within the CI of the Zezelj and Jokic study?
1 (EL&T Study 2, 2008)	0.66 ± 0.27	−0.06 ± 0.37	−0.78 ± 0.59	No	Zezelj and Jokic not, other two significant	No
2 (EL&T Study 3, 2008)	0.71 ± 0.17	0.61 ± 0.14		Yes	Both significant	Yes
3 (EL&T Study 4, 2008)	0.80 ± 0.26	−0.26 ± 0.26		No	Eyal et al. significant, Zezelj and Jokic not	No
4 (G&M Study 1, 2012)	−0.34 ± 0.24	−0.68 ± 0.36	−1.06 ± 0.35	Yes	Eyal et al. not, other two significant	No

Notes. As in some of the original studies samples were relatively small ($n < 20$), and Cohen's d gives a biased estimate of the population effect size especially for small samples, we opted for *corrected effect size*, Hedges's g (recommended in Cumming, 2013; Lakens, 2013).

Following the recommendations from Valentine et al. (2011) and Lakens (2013), we summarized the results of four experiments in Table 2.

Discussion

To set the right tone for this discussion, we must first acknowledge that there is no such thing as an exact replication – there are always known and unknown factors that possibly contribute to a certain outcome that might not be identical between two studies (from characteristics of participants to physical conditions and time of day). We share the view of other social scientists (e.g., Asendorpf et al., 2013; Cacioppo & Cacioppo, 2013; Spellman, 2013) that this fact, however, should not discourage researchers from performing replications, as it is a necessary step for further generalization and/or establishing the limits of a certain phenomenon. In our replication study, we put every effort to have identical stimuli, procedure and participants as in original studies. We were fortunate enough to have the full cooperation from the other two laboratories – they were fast in sharing materials, instruments, and databases upon our request. The only known difference between the studies was the fact that samples were drawn from different cultures. Strictly speaking, this should not be decisive to the main manipulation effect: Cognitive construals should be sensitive to psychological distance, and this mechanism should serve self-regulatory needs, that are culturally invariant. However, EL&T application of CLT was in the domain of moral reasoning, which has proven to be culturally sensitive (e.g., Boyes & Walker, 1988; Snarey, 1985; Tsui & Windsor, 2001). Moreover, there were attempts to attribute current inconclusive results to cultural differences (G&M, 2012), which is why a direct replication of the original experiments in different cultural setting seemed to be an appropriate starting point.

The original study's hypotheses were derived from the premises of Construal Level Theory (Trope & Liberman, 2010). CLT proposes that distant events are represented more abstractly, globally than events that are psychologically closer. Therefore, a distant event should be evaluated in terms of more primary, high-level features. Applied in the domain of moral reasoning, this translates to expectation that one's reliance on universal moral principles and neglect to the detail should be enhanced in evaluating psychologically distant events.

We begun our investigation following two conflicting sets of findings: One supporting the conclusion that high-level construal leads to less sensitivity to context – therefore to harsher judgment of moral transgressions and more appreciation to virtuous behavior (EL&T, 2008), and the other supporting the conclusion that low construal leads to more sensitivity to context and therefore to an opposite moral evaluation of acts (G&M, 2012). Given that the cooperative efforts of the two research groups did not lead

---

7   Effect sizes were calculated using De Fife (2009). Effect size calculator, Emory University. Retrieved from http://www.psychsystems.net/Manuals/

to a resolution, it was necessary to further disentangle this puzzle.

Our four attempts to replicate experiments investigating the impact of psychological distance on moral judgment yielded three different outcomes: No systematic effect of temporal distance (regardless of the nature of the act: Transgression or virtuous), the effect of social distance compatible with CLT predictions, and the reversed effect of direct construal level manipulation.

There is accumulated empirical evidence demonstrating that both temporal and social distance indeed affects moral judgment. For example, research done in Sweden (Agerström & Björklund, 2009; Agerström, Björklund, & Carlsson, 2013) concluded that people make more extreme moral judgments of behavior from a distant than from a near time (or visual) perspective, and that this effect was driven by the level of construal. Lammers (2012) demonstrated in four studies that subjects reacted more negatively to *others*' morally questionable behaviors when they took an abstract (high-level) perspective rather than a concrete (low-level) perspective. However, they were inclined to react less negatively to *their own* moral transgressions.

CLT does not assume different impact of temporal and social distance – in fact, it was initially a theory of temporal construal, exploring the effects of time perspective on mental representation (Liberman, Sagristano, & Trope, 2002; Liberman & Trope, 1998; Trope & Liberman, 2003). Only later it was generalized to other forms of psychological perspectives, namely social, spatial, and hypothetical (Trope & Liberman, 2010). It is therefore difficult to speculate the reasons as to why the temporal distance manipulation showed no consistent impact on morality judgments.

The effects of direct priming of construal level within the "how-and-why" task were in line with the results obtained by G&M. This task was designed to activate a low construal mindset (through a series of "how questions" emphasizing the means by which activities are carried out) or a high construal mindset (through a series of "why questions" emphasizing the end state activities lead to). The priming procedure was developed by Freitas, Gollwitzer, and Trope (2004), and successfully implemented for this purpose in other experiments within CLT framework (e.g., Wakslak & Trope, 2009). In our research it was expected to shift respondents' focus in evaluation of immoral acts from moral universalities to contextual information that attenuated the severity of the act. However, it could be the case that if a person was prompted to think about the means of immoral acts, they tend to focus on the act itself and thus represent it very concretely and vividly. This representation could invoke strong emotional responses. Alternatively, it could be argued that the details provided with intention to mitigate the severity of the act could have been perceived as implausible excuses and therefore dismissed. Both could paradoxically lead to harsher instead of leaner judgments.

However, the obtained pattern of results is not easily attributable to cultural differences or to unexpected consequences of provided contextual information. Had we discovered, for example, that a low construal mindset consistently leads to a harsher judgment (as G&M had),

we could have speculated that attenuating information would not be plausible to our respondents so that focusing on them did not have the expected effect. Yet we had the strongest effect of social distance manipulation, in which the judgment was harsher for a high construal mindset (i.e., judging moral behavior from a third person perspective). What our replication venture seems to show is that there is a complex interplay between (A) domain of judgment (moral judgment seems to be very specific), (B) procedures employed to invoke a specific mindset (direct priming, temporal distance, and social distance manipulation yield different results in morality assessment, maybe because they do not all necessarily lead to focus on moral universalities or context, as expected), and (C) the ethnicity/culture/society that respondents are recruited from.

Future research could benefit from developing clear manipulation checks aiming to assess if different priming techniques really lead to different levels of mental construal. It could also seek to directly compare the effects of different procedures (social, temporal distance, and direct priming) on moral judgments. This in turn could help in establishing the limits of CLT's generalizability as one of the most promising and influential theories in the field.

Note From the Editors

Commentaries and a rejoinder on this paper are available (Eyal, Liberman, & Trope, 2014; Gong & Medin, 2014; Žeželj & Jokić, 2014; doi: 10.1027/1864-9335/a000206).

References

Agerström, J., & Bjorklund, F. (2009). Moral concerns are greater for temporally distant events and are moderated by value strength. *Social Cognition, 27*, 261–282.

Agerström, J., Björklund, F., & Carlsson, R. (2013). Look at yourself! Visual perspective influences moral judgment by level of mental construal. *Social Psychology, 44*, 42–46.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108–119.

Boyes, M. C., & Walker, L. J. (1988). Implications of cultural diversity for the universality claims of Kohlberg's theory of moral reasoning. *Human Development, 31*, 44–59.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216.

Brislin, R. W. (1976). Comparative research methodology: Cross-cultural studies. *International Journal of Psychology, 11*, 215–229.

Cacioppo, J.T., & Cacioppo, S. (2013). Minimal replicability, generalizability and scientific advances in psychological science. *European Journal of Personality, 27*, 121–122.

Cumming, G. (2013). The new statistics: A how-to guide. *Australian Psychologist, 48*, 161–170.

Eyal, T., Liberman, N., & Trope, Y. (2008). Judging near and distant virtue and vice. *Journal of Experimental Social Psychology, 44*, 1204–1209.

Eyal, T., Liberman, N., & Trope, Y. (2014). A comment on Žeželj and Jokić replication. Commentaries and rejoinder to Žeželj and Jokić (2014). *Social Psychology*. Advance online publication. doi: 10.1027/1864-9335/a000206.

Freitas, A., Gollwitzer, P., & Trope, Y. (2004). The influence of abstract and concrete mindsets on anticipating and guiding others' self-regulatory efforts. *Journal of Experimental Social Psychology, 40*, 739–752.

Gong, H., & Medin, D.L. (2012). Construal levels and moral judgment: Some complications. *Judgment and Decision Making, 7*, 628–638.

Gong, H., & Medin, D. L. (2014). Commentary on Žeželj and Jokić (2014). Commentaries and rejoinder to Žeželj and Jokić (2014). *Social Psychology*. Advance online publication. doi: 10.1027/1864-9335/a000206

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 863–900.

Lammers, J. (2012). Abstraction increases hypocrisy. *Journal of Experimental Social Psychology, 48*, 475–480.

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.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology, 38*, 523–534.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Shapira, O., Liberman, N., Trope, N., & Rim, S. (2012). Levels of mental construal. In S. Fiske & N. McRae (Eds.), *Sage handbook of Social Cognition* (pp. 229–250). London: Sage.

Snarey, J. R. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin, 97*, 202.

Spellman, B. A. (2013). There is no such thing as replication, but we should do it anyway. *European Journal of Personality, 27*, 136–137.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*, 403–421.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*, 440–463.

Trope, Y., & Liberman, N. (2011). Construal level theory. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 118–135). London, UK: Sage.

Tsui, J., & Windsor, C. (2001). Some cross-cultural evidence on ethical reasoning. *Journal of Business Ethics, 31*, 143–150.

Valentine, J. C., Biglan, A., Boruch, R. F., Castro, F. G., Collins, L. M., Flay, B.R., . . . Schinke, S. P. (2011). Replication in prevention science. *Prevention Science, 12*, 103–117.

Wakslak, C., & Trope, Y. (2009). The effect of construal level on subjective probability estimates. *Psychological Science, 20*, 52–58.

Žeželj, I. L., & Jokić, B. R. (2014). A rejoinder to comments by Eyal, Liberman & Trope and Gong & Medin. Commentaries and rejoinder to Zezelj and Jokic (2014). *Social Psychology*. Advance online publication. doi: 10.1027/1864-9335/a000206

Iris Žeželj

Faculty of Philosophy
University of Belgrade
Čika Ljubina 18-20
11000 Belgrade
Serbia
E-mail izezelj@f.bg.ac.rs