Supplementary Materials: Distancing and Moral Dumbfounding Pilot Studies

Cillian McHugh[1], Marek McGann[2], Eric R. Igou[1], and Elaine L. Kinsella [1]

[1]Department of Psychology, University of Limerick

[2]Department of Psychology, Mary Immaculate College

Author Note

All procedures performed in studies involving human participants were approved by the Institutional Research Ethics Committee and conducted in accordance with the Code of Professional Ethics of the Psychological Society of Ireland and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Correspondence concerning this article should be addressed to Cillian McHugh, Email: cillian.mchugh@ul.ie

Abstract

Moral dumbfounding occurs when people maintain a moral judgment in the absence of supporting reasons. Drawing on dual-process approaches to moral judgment, one possible explanation for moral dumbfounding proposes that it occurs as a result of a conflict between intuitive and deliberative processes. Consistent with this explanation, previous research has shown that under manipulations designed to lead to more intuitive thinking rather than deliberative thinking (such as increased cognitive load), people are less likely to provide reasons for their judgments, and more likely to provide dumbfounded responses in a moral dumbfounding task. Building on this work the present research examines if dumbfounded responding can be reduced through experimental manipulations designed to facilitate deliberative thinking (over intuitive thinking). Drawing on construal-level theory, and the finding that distancing facilitates deliberative thinking, we predict that including a distancing manipulation in a moral dumbfounding task will increase reason-giving, and reduce dumbfounded responding. We propose a pre-registered study to test this prediction.

Keywords: moral dumbfounding, distancing, construal-level theory, dual-processes, reasons, intuitions

Supplementary Materials: Distancing and Moral Dumbfounding Pilot Studies

Below we present eight pilot studies to test the hypothesized relationship between distancing and dumbfounded responding. Recent work (McHugh et al., 2023) has proposed a conflict in dual-processes (e.g., Bonner & Newell, 2010) explanation of moral dumbfounding. According to this view, dumbfounding occurs when a habitual response (moral judgment) is in conflict with a deliberative response (providing reasons). This explanation is consistent with dual-process approaches to moral judgment (e.g., Bago & De Neys, 2019; Cushman, 2013; Greene, 2008), as well as with a unimodel (Kruglanski & Gigerenzer, 2011) and categorization (McHugh et al., 2022) approaches.

A key prediction of this explanation is that rates of reason-giving should be influenced by experimental manipulations that impact intuitive vs deliberative thinking. Previous work has demonstrated that inhibiting deliberative thinking through a cognitive load manipulation can reduce reason-giving, leading to higher rates of dumbfounding (McHugh et al., 2023). A corollary of this finding is that reason-giving should be increased under manipulations that encourage deliberative thinking. Drawing on construal-level theory (Förster et al., 2004; Liberman et al., 2002), we predict that increased distance will facilitate the identification of reasons, leading to lower levels of dumbfounded responding.

The Current Pilot Studies

We tested the feasibility of different manipulations of both temporal distance and social distance. A-priori power analysis indicated that in order to detect a large effect size ($V = .35$) with 80% power, a sample of $N = 79$ participants is required, in order to detect a medium effect size ($V = .21$) with 80% power a sample of $N = 218$ participants is required, and in order to detect, a small effect size ($V = .\ 07$) with 80% power a sample of $N = 1966$ is required. The pilot studies below are sufficiently powered to detect large effects only. Based on previous research investigating influences on moral dumbfounding we anticipate a small to medium effect size (e.g., McHugh et al., 2023, report $V = 0.12$). Despite being under-powered, these studies are

descriptively informative, and provide a preliminary estimate of the direction of any possible effect that would be expected from a higher powered study.

Pilot Studies 1 and 2 employ a manipulation of social distance compared against a control condition (no manipulation). Pilot Studies 3, 4, 6, and 7 employ a temporal distance manipulation, comparing increased temporal distance against decreased temporal distance. For consistency across conditions, and in order to enhance the plausibility of the materials, participants in these studies were encouraged to think about the scenarios from the perspective of a third person (thus these studies additionally included a social distance manipulation, however this was kept constant across conditions in these studies). Pilot Studies 5 and 8 were also employed a temporal distance manipulation, and participants were encouraged to think about the scenarios from a first person perspective. Pilot Studies 1-5, and 8 used the *Julie and Mark* scenario while Pilot Studies 6 and 7 used the *Jennifer* scenario. Pilot Studies 1, 2, 3, 4, 5, 7, and 8, recorded dumbfounded responding using the 'critical slide' (McHugh et al., 2017, 2023), while Pilot Study 6 trialed a more continuous measure of dumbfounded responding. The designs and results of Pilot Studies 1-8 are summarized in Figure 1.

Figure 1

Overview of Pilot Studies 1-8.

Study	Distancing	Manipulation	Conditions	Perspective	Scenario	Measure	N	Result	*p*	direction
Pilot 1	Social	Anne vignette	Present vs Absent	First person	Julie & Mark	Critical Slide	120	No sig. Effect	.292	(predicted)
Pilot 2	Social	Anne vignette	Present vs Absent	First person vs third person	Julie & Mark	Critical Slide	104 / 76	No sig. Effect	.098 / .101	(predicted)
Pilot 3	Temporal	Anne vignette	Increased vs decreased	Third person	Julie & Mark	Critical Slide	117 / 71	Sig. Effect	.010 / .017	predicted
Pilot 4	Temporal	Anne vignette	Increased vs decreased	Third person	Julie & Mark	Critical Slide	165 / 77	No sig. Effect	.741 / .879	(predicted)
Pilot 5	Temporal	Imagined assignment	Increased vs decreased	First Person	Julie & Mark	Critical Slide	97 / 71	No sig. Effect	.432 / .464	(predicted)
Pilot 6	Temporal	Anne vignette	Increased vs decreased	Third person	Jennifer	Scale Measure	158 / 76	No sig. Effect	.306 / .278	(mixed)
Pilot 7	Temporal	Anne vignette	Increased vs decreased	Third person	Jennifer	Critical Slide	159 / 87	No sig. Effect	.868 / .472	(predicted)
Pilot 8	Temporal	Modified Scenario	Increased vs absent	First Person	Julie & Mark	Critical Slide	286 / 124	No sig. Effect	.482 / .681	(opposite)

Note. Sample sizes listed are the total samples that passed the attention checks, without exclusion based on responses to Need for Closure, followed by the sample size when excluding participants who failed the Need for Closure attention check.

Pilot Study 1 - Social Distance, "Julie and Mark" (First Person)

The aim of Pilot Study 1 was to investigate if a social distance manipulation influenced participants' ability to justify their moral judgment. We also measured social desirability (Ballard, 1992; Crowne & Marlowe, 1960; Strahan & Gerbasi, 1972) as a potential correlate/moderator variable.

Pilot Study 1: Method

Pilot Study 1: Participants and Design

Pilot Study 1 was a between-subjects design. The dependent variable was rates of providing reasons/dumbfounding (measured using the critical slide with 3 response options: 1: reason-giving; 2: nothing-wrong; 3: dumbfounding - an admission of not having reasons). The independent variable was social distancing with two levels: present and absent.

To manipulate distancing participants were told that a philosophy student (Anne) had been asked to consider the moral scenario, and participants were asked to consider the reasons they might use to justify Anne's judgement. Social desirability (Ballard, 1992; Crowne & Marlowe, 1960; Strahan & Gerbasi, 1972) was included as an additional potential predictor variable.

A total sample of 120 participants (62 female, 58 male; $M_{\text{age}} = 38$, min = 22, max = 75, $SD = 11.9$) took part. Participants were recruited through MTurk. Participation was voluntary and participants were paid 0.50 US dollars for their participation. Participants were recruited from English speaking countries or from countries where residents generally have a high level of English (e.g., The Netherlands, Denmark, Sweden).

Pilot Study 1: Procedure and Materials

Data were collected using an online survey generated using Questback (Unipark, 2013). The survey opened with questions relating to basic demographics. Following this, participants were presented with two statements relating to the norm principle

(taken from McHugh et al., 2020; Royzman et al., 2015).[1] At this point, the distancing group were presented with an additional set of instructions prior to being presented with the the *Julie and Mark* scenario (taken from Haidt et al., 2000). Participants in the control group did not receive the additional instructions, and proceeded directly to the *Julie and Mark* scenario. The distancing instructions, and the *Julie and Mark* scenario read as follows:

Distancing Instructions.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics and has been asked to study the following scenario.

How should Anne judge the actions of the two people involved?

What reasons would you use to explain why she should make that judgement?

Julie and Mark Scenario.

Julie and Mark, who are brother and sister, are travelling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other.

[1] Responses to this are not relevant for the research question of interest here and as such we do not discuss this in our analyses.

Participants rated how right or wrong the behavior of Julie and Mark was on a 7-point Likert scale (where, 1 = *Morally wrong*; 4 = *neutral*; 7 = *Morally right*), and were asked to provide reasons for their judgement (or what reasons they would use to explain to Anne why she should make that judgment). Participants then read a series of counter-arguments (developed by McHugh et al., 2017), which refuted commonly used justifications for rating the behavior as "wrong".

To measure dumbfounding we used the *critical slide* developed by McHugh et al. (2017). This included a statement defending the behavior and a question asking how the behavior could be wrong ("Julie and Mark's behaviour did not harm anyone, how can there be anything wrong with what they did?"). There were three possible answer options: (a) "It's wrong and I can provide a valid reason" (reason-giving), (b) "It's wrong but I can't think of a reason" (dumbfounding), and (c) "There is nothing wrong" (nothing-wrong). The order of these response options was randomized. The selecting of option (b), the admission of not having reasons, was taken to be a dumbfounded response, and we note that this measure provides a conservative measure of dumbfounded responding (McHugh et al., 2017). Participants who selected (a) were prompted to type a reason once they progressed to the next page.

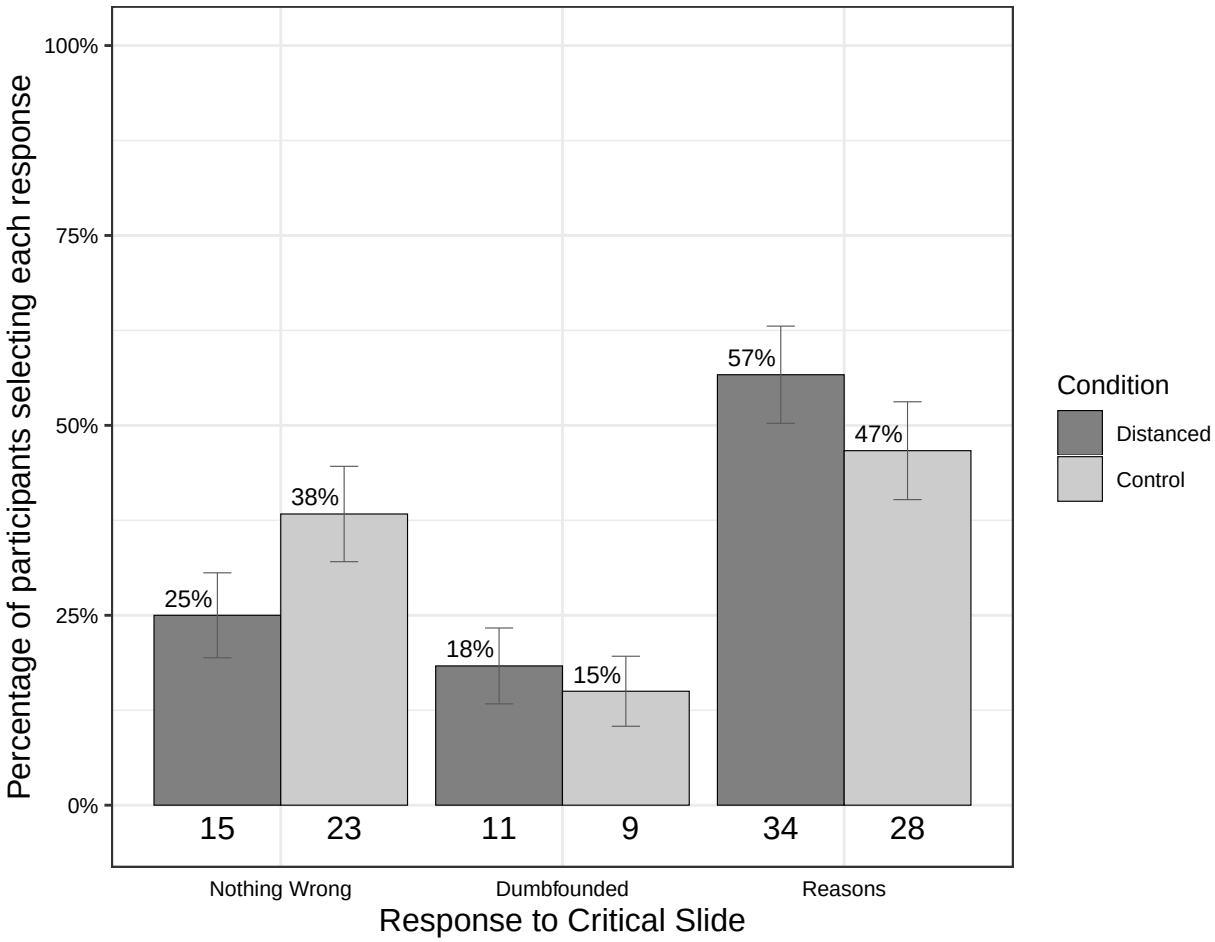
Following the critical slide rated the behavior again on a 7-point Likert scale (where, 1 = *Morally wrong*; 4 = *neutral*; 7 = *Morally right*). They then responded to the credulity check questions devised by Royzman et al. (2015), and answered the three questions relating to the application of the harm principle (McHugh et al., 2020).[2] Finally participants completed the short version of the Marlowe-Crowne (Crowne & Marlowe, 1960) social desirability scale (devised by Strahan & Gerbasi, 1972; see also Ballard, 1992). This consisted of ten questions (e.g., "There have been occasions when I took advantage of someone.", "I never resent being asked to return a favor.") to which participants selected "True" or "False".

[2] As with the norm principles questions above responses to these questions are not relevant for the research question of interest here and as such we do not discuss this in our analyses.

Pilot Study 1: Results

Figure 2

Study 1: Responses to critical slide for the experimental group (N = 60) and the control group (N = 60); (error bars represent standard error of the proportion)



Overview of Judgments

Seventy six participants (63.87%) rated the behavior of Julie and Mark as wrong initially, and Seventy two participants (60.5%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.9, SD = 2.1) and revised ratings (M = 2.9, SD = 2), $t(118)$ = -0.72, p = .476, d = 0.07.

Distancing and Judgments Made

There was no difference in initial judgement depending on distance manipulation: $t(113.8)$ = -0.93, p = .356, d = 0.17, ($M_{\text{distanced}}$ = 2.7, $SD_{\text{distanced}}$ = 2.2, M_{control} = 3.1, SD_{control} = 1.9). There was no difference in revised judgement depending

on distance manipulation: $t(113.99) = -1.23$, $p = .223$, $d = 0.22$, ($M_{\text{distanced}} = 2.7$, $SD_{\text{distanced}} = 2.2$, $M_{\text{control}} = 3.2$, $SD_{\text{control}} = 1.9$).

Distancing and Reason-Giving/Dumbfounding

There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 119) = 2.286$, $p = .319$, $V = 0.14$, the observed power was 0.25. The responses to the critical slide for the experimental group ($N = 59$) and the control group ($N = 60$) are displayed in Figure 2.

Social Desirability and Dumbfounding

There was no significant association between Social Desirability and response to the critical slide $\chi^2(2, N = 119) = 0.84$, $p = .656$, The observed power was 0.12.

Pilot Study 2 - Social Distance, "Julie and Mark" (First Person vs Third Person)

As with Pilot Study 1, the aim of Pilot Study 2 was to investigate if a social distance manipulation influenced participants' ability to justify their moral judgment. There were two changes from Pilot Study 1. First, we modified our distancing manipulation to include an explicit instruction to think about the scenario from the perspective of a third party. In addition to recording social desirability (Ballard, 1992; Crowne & Marlowe, 1960; Strahan & Gerbasi, 1972), Pilot Study 2 additionally included need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and the short form of the cognitive reflection test (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) as potential correlate/moderator variables.

Pilot Study 2: Method

Pilot Study 2: Participants and Design

Pilot Study 2 was a between-subjects design. The dependent variable was rates of providing reasons/dumbfounding - measured using the critical slide (with the same 3 response options: 1: reason-giving; 2: nothing-wrong; 3: dumbfounding - an admission of not having reasons). The independent variable was social distancing with two levels: present and absent. The distancing manipulation was similar to Pilot Study 1 (the *Anne* vignette) with the inclusion of an explicit instruction to consider the scenario from Anne's perspective. Social desirability (Ballard, 1992; Crowne & Marlowe, 1960; Strahan & Gerbasi, 1972), need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

A total sample of 104 participants (49 female, 55 male; $M_{\text{age}} = 37.5$, min = 19, max = 83, $SD = 12.2$) took part. The measure of need for closure includes a "lie score", whereby if participants score above a threshold on a combination of specific items they are deemed to be lying (example lie score items include claiming to never have been late for an appointment, or never having met someone the didn't like). Following the removal of participants who scored above the lie score threshold, we were left with a

sample of 76 participants (37 female, 39 male; $M_{\text{age}} = 38.7$, min = 19, max = 83, $SD = 12.6$) who were eligible for analysis. Participants were recruited through MTurk in the same way as in Pilot Study 1 (same payment amount, same country selection).

Pilot Study 2: Procedure and Materials

The procedure and materials were similar to Pilot Study 1 with a change to the distancing manipulation and the inclusion of additional measures. The distancing manipulation for Pilot Study 2 included an explicit instruction to think about the scenario from the perspective of a third party. The revised manipulation read as follows:

Distancing Instructions.

> Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she adopts a position on an issue in class and struggles (or fails) to defend it when challenged by others.
>
> She is currently taking a course in ethics and has been asked to study the following scenario.
>
> While reading the story on the next page, try to imagine how the philosophy student Anne will judge the actions of the two people.
>
> In particular try to think about reasons she may use to defend her judgement.
>
> Try to think about the story from Anne's perspective rather than your own.

In addition to social desirability, we also measured need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) cognitive reflection (CRT Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011). The need for closure scale contains 47 questions (e.g., "I'd rather know bad news than stay in a state of uncertainty.") to which participants respond on a 6 point Likert scale, where 1 = *strongly disagree*, and 6 = *strongly agree*. The CRT is a brief test of analytical thinking. It contains three

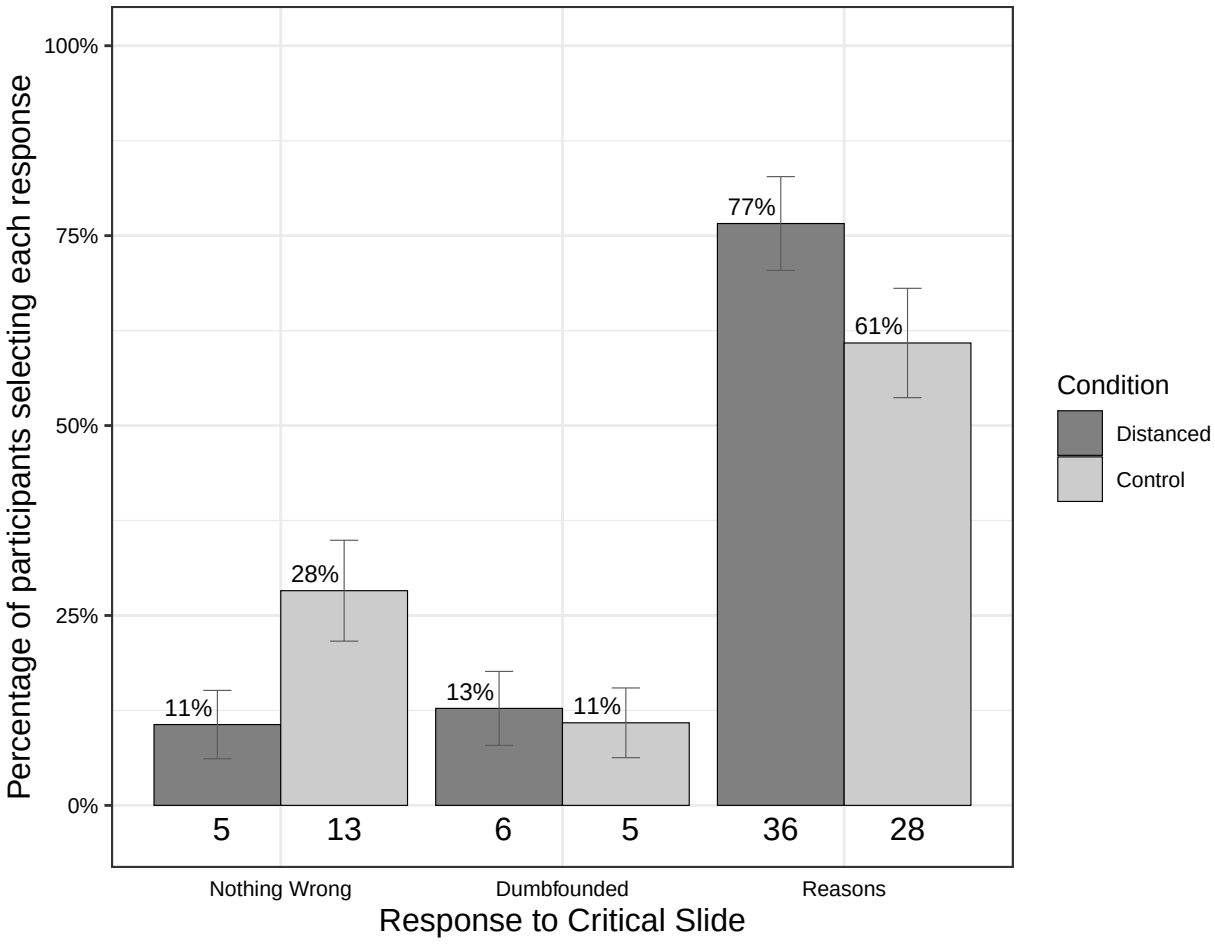
questions, each of which has an answer that seems intuitively correct, but is actually wrong (e.g., If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?)

Pilot Study 2: Results

Below we present two sets of results. First we present the results for the full sample, second we present the results for the sample with exclusions based on participants' "lie score".

Figure 3

Pilot Study 2: Responses to critical slide for the experimental group (N = 47) and the control group (N = 46); (No exclusions; error bars represent standard error of the proportion)



No Exclusions

Overview of Judgments (no exclusions). Seventy two participants (69.23%) rated the behavior of Julie and Mark as wrong initially, and Sixty six participants (63.46%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.5, SD = 2) and revised ratings (M = 2.4, SD = 2), $t(89) = 0$, $p = 1.000$, $d = 0$.

Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(96.67) = 0.35$, $p = .729$, $d = 0.07$, ($M_{\text{distanced}} = 2.5$, $SD_{\text{distanced}} = 2$, $M_{\text{control}} = 2.4$, $SD_{\text{control}} = 2.1$). There was no difference in revised judgement depending on distance manipulation: $t(86.71) = -0.27$, $p = .789$, $d = 0.06$, ($M_{\text{distanced}} = 2.3$, $SD_{\text{distanced}} = 1.9$, $M_{\text{control}} = 2.5$, $SD_{\text{control}} = 2.2$).

Distancing and Reason-Giving/Dumbfounding (no exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 104) = 4.636$, $p = .098$, $V = 0.21$, the observed power was 0.47. The responses to the critical slide for the experimental group ($N = 52$) and the control group ($N = 52$) are displayed in Figure 3.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(6, N = 104) = 2.8$, $p = .834$, The observed power was 0.3.

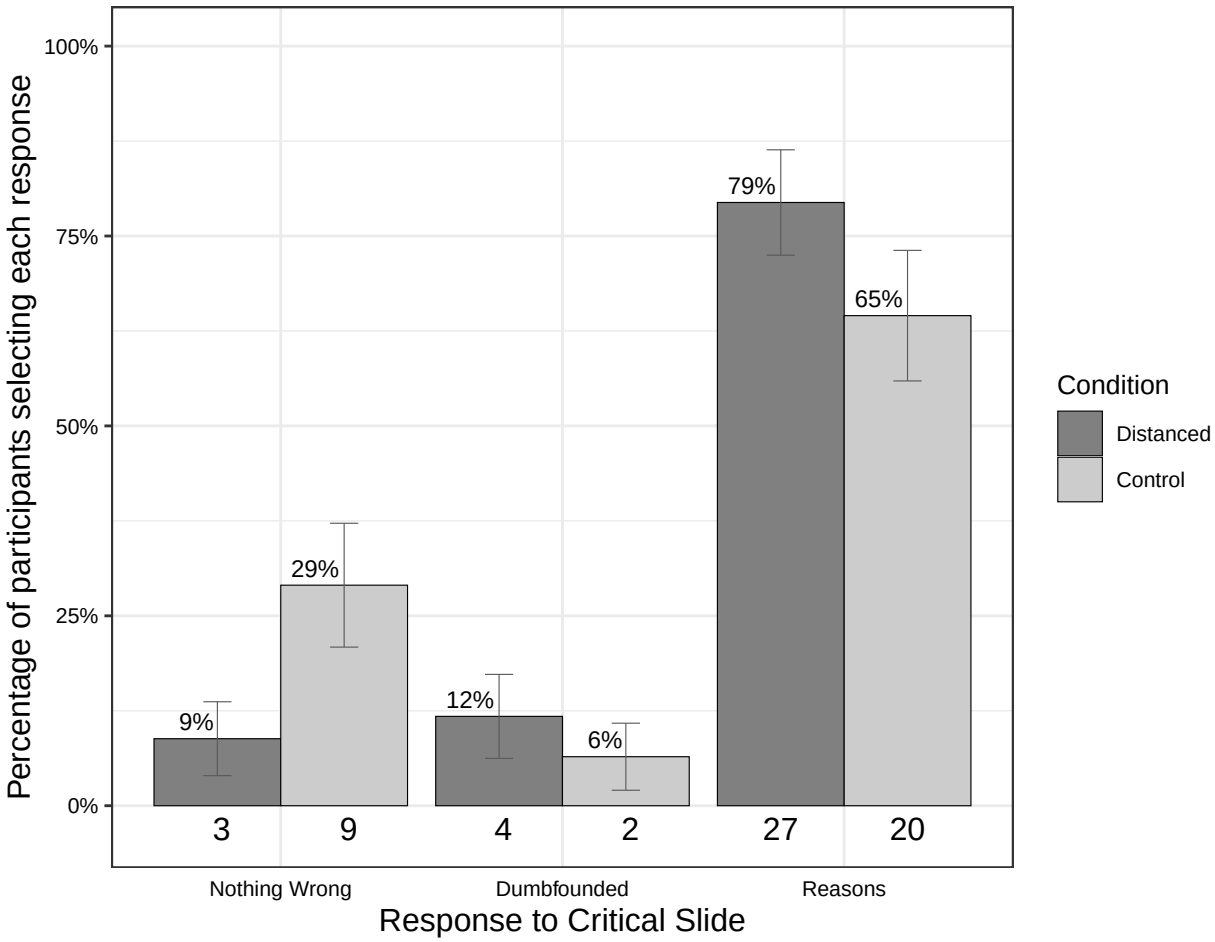
With Exclusions

Overview of Judgments (with exclusions). Fifty one participants (67.11%) rated the behavior of Julie and Mark as wrong initially, and Forty seven participants (61.84%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.5, SD = 2.1) and revised ratings (M = 2.4, SD = 2.1), $t(61) = 0.65$, $p = .517$, $d = 0.08$.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(66.79) = 0.22$, $p = .827$, $d = 0.05$, ($M_{\text{distanced}} = 2.6$, $SD_{\text{distanced}} = 2$, $M_{\text{control}} = 2.5$, $SD_{\text{control}} = 2.2$).

Figure 4

Pilot Study 2: Responses to critical slide for the experimental group (N = 34) and the control group (N = 31); (With exclusions; error bars represent standard error of the proportion)



There was no difference in revised judgement depending on distance manipulation: $t(56.44) = -0.61$, $p = .543$, $d = 0.16$, ($M_{\text{distanced}} = 2.2$, $SD_{\text{distanced}} = 1.8$, $M_{\text{control}} = 2.5$, $SD_{\text{control}} = 2.3$).

Distancing and Reason-Giving/Dumbfounding (with exclusions).

There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 76) = 4.581$, $p = .101$, $V = 0.25$, the observed power was 0.47. The responses to the critical slide for the experimental group ($N = 39$) and the control group ($N = 37$) are displayed in Figure 4.

Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(6, N = 76) = 5.36$, $p = .499$, The observed power was 0.53.

Pilot Study 3 - Temporal Distance, "Julie and Mark" (Third Person)

For both Pilot Studies 1 and 2 there was no significant association between response to the critical slide and the social distance manipulation. Pilot Study 3 was designed to test an alternative distance manipulation, temporal distance. The aim of Pilot Study 3 was to investigate if manipulating temporal distance influenced participants' ability to justify their moral judgment.

Pilot Study 3: Method

Pilot Study 3: Participants and Design

Pilot Study 3 was a between-subjects design. The dependent variable was rates of reason-giving/dumbfounding - measured using the critical slide. The independent variable was temporal distance with two levels: increased and decreased. All participants read a vignette similar to the distancing manipulation used in Pilot Studies 1 and 2 and temporal distance was manipulated by varying the deadline by which the assignment would need to be completed by - either five weeks in the future (increased) or today (decreased). Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables. Social desirability was not recorded.

A total sample of 105 participants (47 female, 58 male; $M_{\text{age}} = 37.5$, min = 21, max = 73, $SD = 12.5$) took part. As in Pilot Study 2, participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 71 participants (35 female, 36 male; $M_{\text{age}} = 40.7$, min = 21, max = 73, $SD = 13.4$). Participants were recruited through MTurk in the same way as in Pilot Study 1 (same payment amount, same country selection).

Pilot Study 3: Procedure and Materials

The procedure and materials were similar to Pilot Study 2 with a change to the distance manipulation and the removal of the social desirability measure. The same moral scenario ("Julie and Mark") was used. The revised manipulation read as follows:

Increased Temporal Distance.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her judgement. **She must complete and submit this assignment within the next five weeks**.

While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Decreased Temporal Distance:.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her judgement. **She must complete and submit this assignment today**.

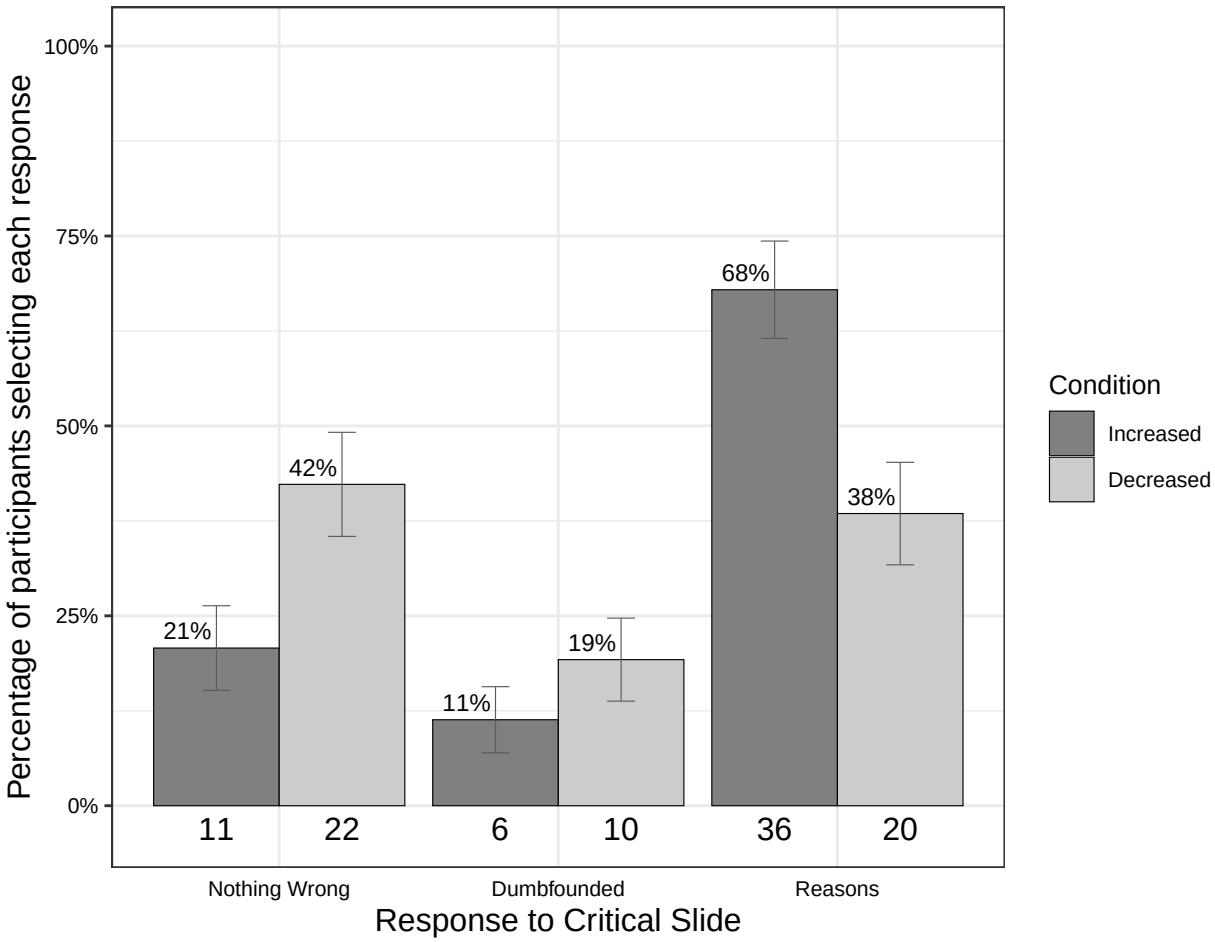
While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Pilot Study 3: Results

As for Pilot Study 2, we present two sets of results. First we present the results for the full sample, second we present the results for the sample with exclusions based on participants' "lie score".

Figure 5

Pilot Study 3: Responses to critical slide for the increased temporal distance group (N = 53) and the reduced temporal distance group (N = 52); (No exclusions; error bars represent standard error of the proportion)



No Exclusions

Overview of Judgments (no exclusions). Sixty three participants (60%) rated the behavior of Julie and Mark as wrong initially, and Sixty four participants (60.95%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.8, SD = 2) and revised ratings (M = 2.9, SD = 2), $t(104)$ = -0.58, p = .566, d = 0.06.

Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(102.72)$ = -1, p = .318, d = 0.2, ($M_{\text{increased}}$ = 2.6, $SD_{\text{increased}}$ = 2, $M_{\text{decreased}}$ = 3, $SD_{\text{decreased}}$ = 2.1). There was no

Table 1

Pilot Suty 3 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on distancing (no exclusions)

		Increased	Decreased
Observed count	Nothing Wrong	11	22
	Dumbfounded	6	10
	Reasons	36	20
Expected count	Nothing Wrong	16.66	16.34
	Dumbfounded	8.08	7.92
	Reasons	28.27	27.73
Standardised residuals	Nothing Wrong	-2.38*	2.38*
	Dumbfounded	-1.13	1.13
	Reasons	3.03*	-3.03*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

difference in revised judgement depending on distance manipulation: $t(102.97) = -1.32$, $p = .190$, $d = 0.26$, ($M_{\text{increased}} = 2.6$, $SD_{\text{increased}} = 2$, $M_{\text{decreased}} = 3.2$, $SD_{\text{decreased}} = 2$).

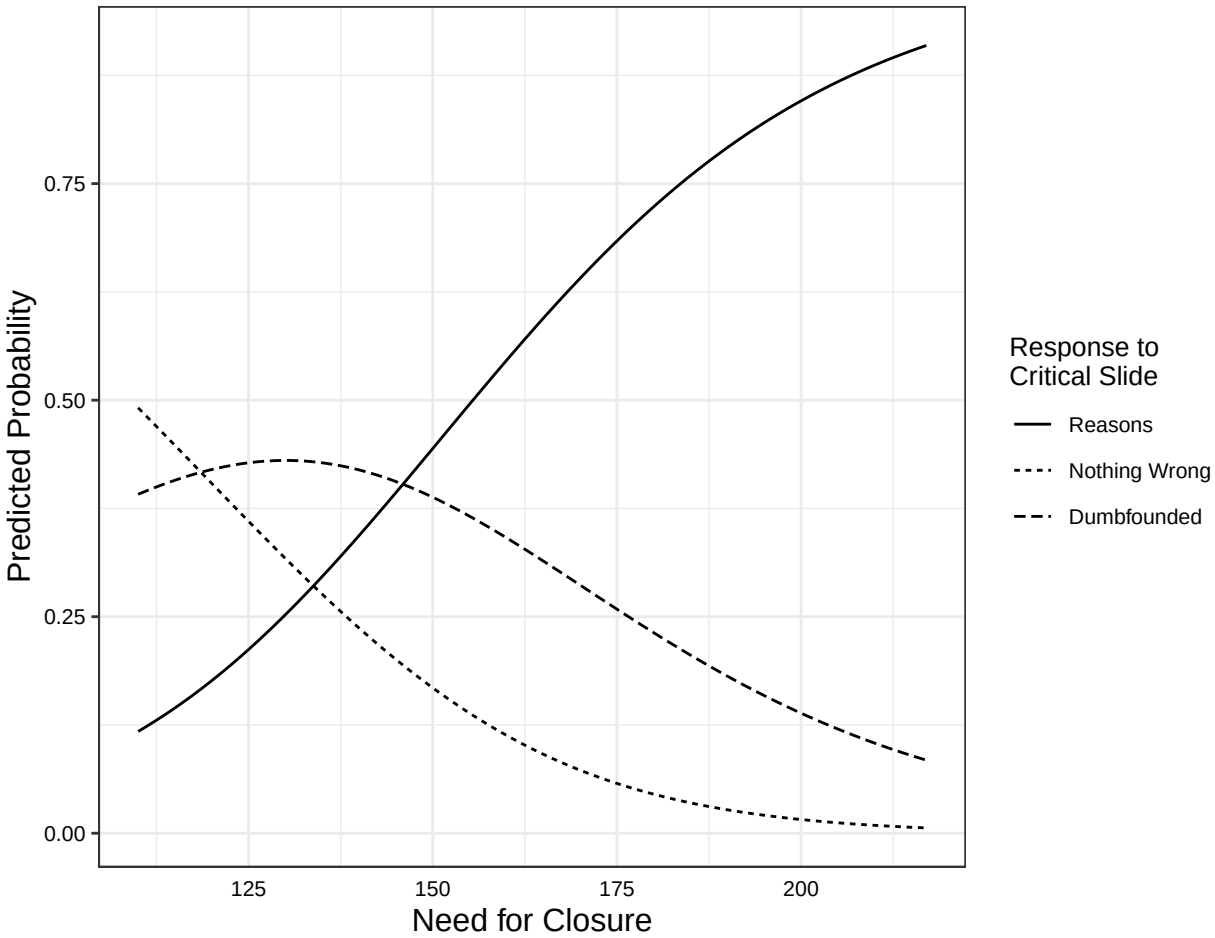
Distancing and Reason-Giving/Dumbfounding (no exclusions). There was a significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 105) = 9.229$, $p = .010$, $V = 0.3$, the observed power was 0.78. The responses to the critical slide for the experimental group ($N = 53$) and the control group ($N = 52$) are displayed in Figure 5. The observed counts, expected counts and standardized residuals are displayed in Table 1.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model significantly predicted responses to the critical slide $\chi^2(4, N = 105) = 13.22$, $p = .010$, The observed power was 0.91. The model explained

between 4.09% (Cox and Snell R square) and 5.75% (Nadelkerke R squared) of the variance in responses to the critical slide. Need for closure was a significant predictor in the model. As need for closure increased, participants were significantly more likely to provide reasons than to present as dumbfounded, Wald = 7.25, p = .007, odds ratio = 0.94, 95% CI [0.9, 0.98], or to select nothing-wrong, Wald = 5.66, p = .017, odds ratio = 0.97, 95% CI [0.94, 0.99]. The predicted probabilities of each response depending on need for closure are displayed in Figure 6.

Figure 6

Study 3: Probability of selecting each response to the critical slide depending on Social Desirability (no exclusions)

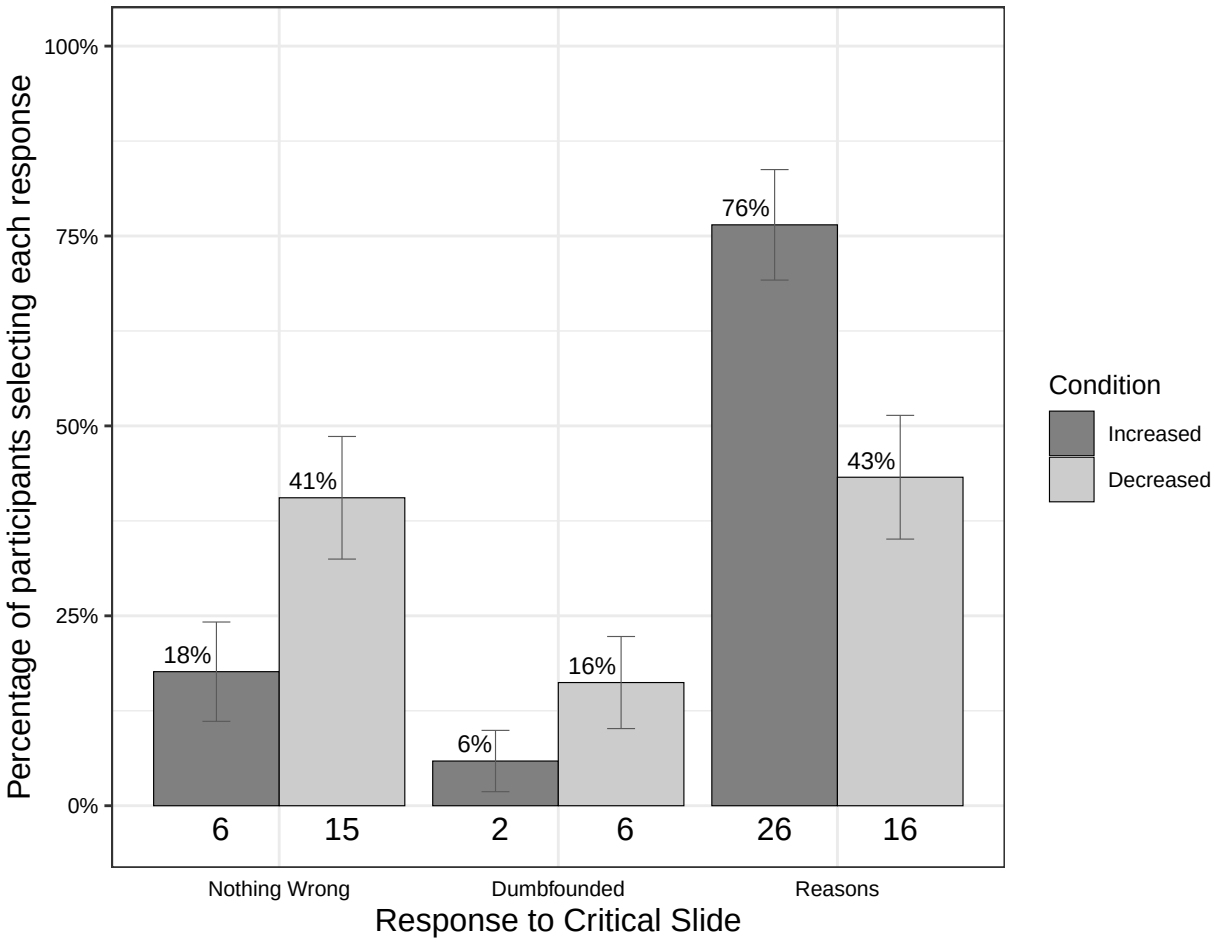


With Exclusions

Overview of Judgments (with exclusions). Fifty participants (70.42%) rated the behavior of Julie and Mark as wrong initially, and Forty nine participants

Figure 7

Pilot Study 3: Responses to critical slide for the increased temporal distance group (N = 34) and the reduced temporal distance group (N = 37); (With exclusions; error bars represent standard error of the proportion)



(69.01%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.6, SD = 2) and revised ratings (M = 2.7, SD = 2), $t(70)$ = -0.83, p = .410, d = 0.1.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(68.71)$ = -1.63, p = .107, d = 0.39, ($M_{\text{distanced}}$ = 2.2, $SD_{\text{distanced}}$ = 1.8, $M_{\text{decreased}}$ = 2.9, $SD_{\text{decreased}}$ = 2.1). Revised judgments were significantly more severe in the increased temporal distance condition ($M_{\text{increased}}$ = 2.2, $SD_{\text{increased}}$ = 1.9) than in the reduced temporal distance condition ($M_{\text{decreased}}$ = 3.1, $SD_{\text{decreased}}$ = 2.1), $t(69)$ = -2.03, p = .046, d = 0.48,

Table 2

Pilot Study 3 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on distancing (with exclusions)

		Increased	Decreased
Observed count	Nothing Wrong	6	15
	Dumbfounded	2	6
	Reasons	26	16
Expected count	Nothing Wrong	10.06	10.94
	Dumbfounded	3.83	4.17
	Reasons	20.11	21.89
Standardised residuals	Nothing Wrong	-2.11*	2.11*
	Dumbfounded	-1.38	1.38
	Reasons	2.85*	-2.85*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Distancing and Reason-Giving/Dumbfounding (with exclusions). There was a significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 71) = 8.126$, $p = .017$, $V = 0.34$, the observed power was 0.72. The responses to the critical slide for the experimental group ($N = 34$) and the control group ($N = 37$) are displayed in Figure 7. The observed counts, expected counts and standardized residuals are displayed in Table 2.

Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4, N = 71) = 6.77$, $p = .149$, The observed power was 0.64.

Pilot Study 4 - Temporal Distance, "Julie and Mark" (Third Person)

Pilot Study 4 was an attempted replication of Pilot Study 3. The aim of Pilot Study 4 was to investigate if temporal distance influenced participants ability to justify moral judgments.

Pilot Study 4: Method

Pilot Study 4: Participants and Design

Pilot Study 4 was an attempted replication of Pilot Study 3. Pilot Study 4 was a between-subjects design. The dependent variable was rates of reason-giving/dumbfounding (measured using the critical slide). The independent variable was temporal distance with two levels, increased and decreased. Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

A total sample of 142 participants (67 female, 74 male; $M_{\text{age}} = 36.2$, min = 19, max = 71, $SD = 11.8$) took part. Again, participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 77 participants (42 female, 35 male; $M_{\text{age}} = 39.3$, min = 22, max = 69, $SD = 13$). Participants were recruited through MTurk in the same way as in previous studies (same payment amount, same country selection).

Pilot Study 4: Procedure and Materials

The procedure and materials were the same as Pilot Study 3.

Pilot Study 4: Results

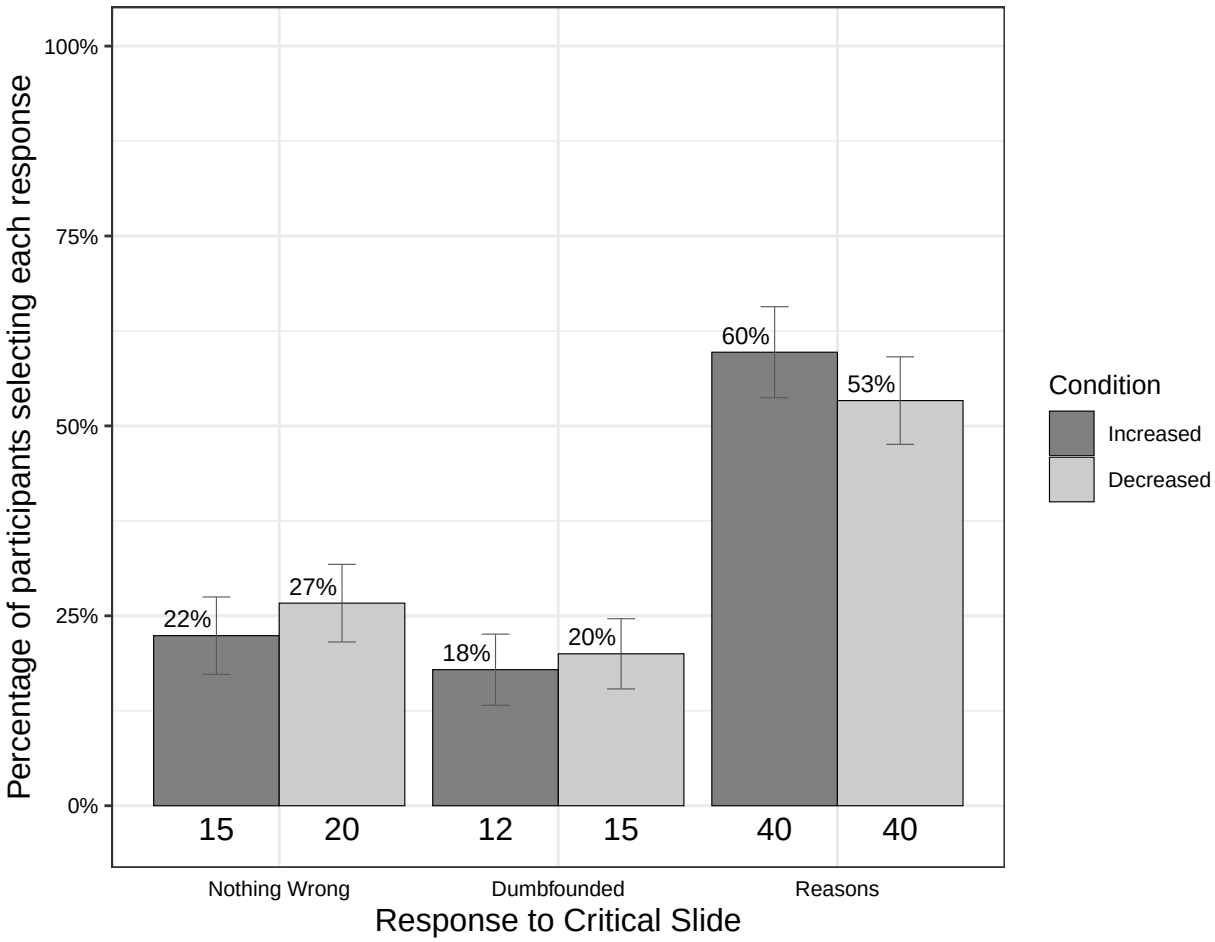
Again, we present two sets of results. First we present the results for the full sample, second we present the results for the sample with exclusions based on participants' "lie score".

No Exclusions

Overview of Judgments (no exclusions). Eighty seven participants (61.3%) rated the behavior of Julie and Mark as wrong initially, and Eighty four participants

Figure 8

Pilot Study 4: Responses to critical slide for the increased temporal distance group (N = 67) and the reduced temporal distance group (N = 75); (No exclusions; error bars represent standard error of the proportion)



(59.2%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.9, SD = 2.2) and revised ratings (M = 3, SD = 2.1), $t(141)$ = -1.2, p = .231, d = 0.1.

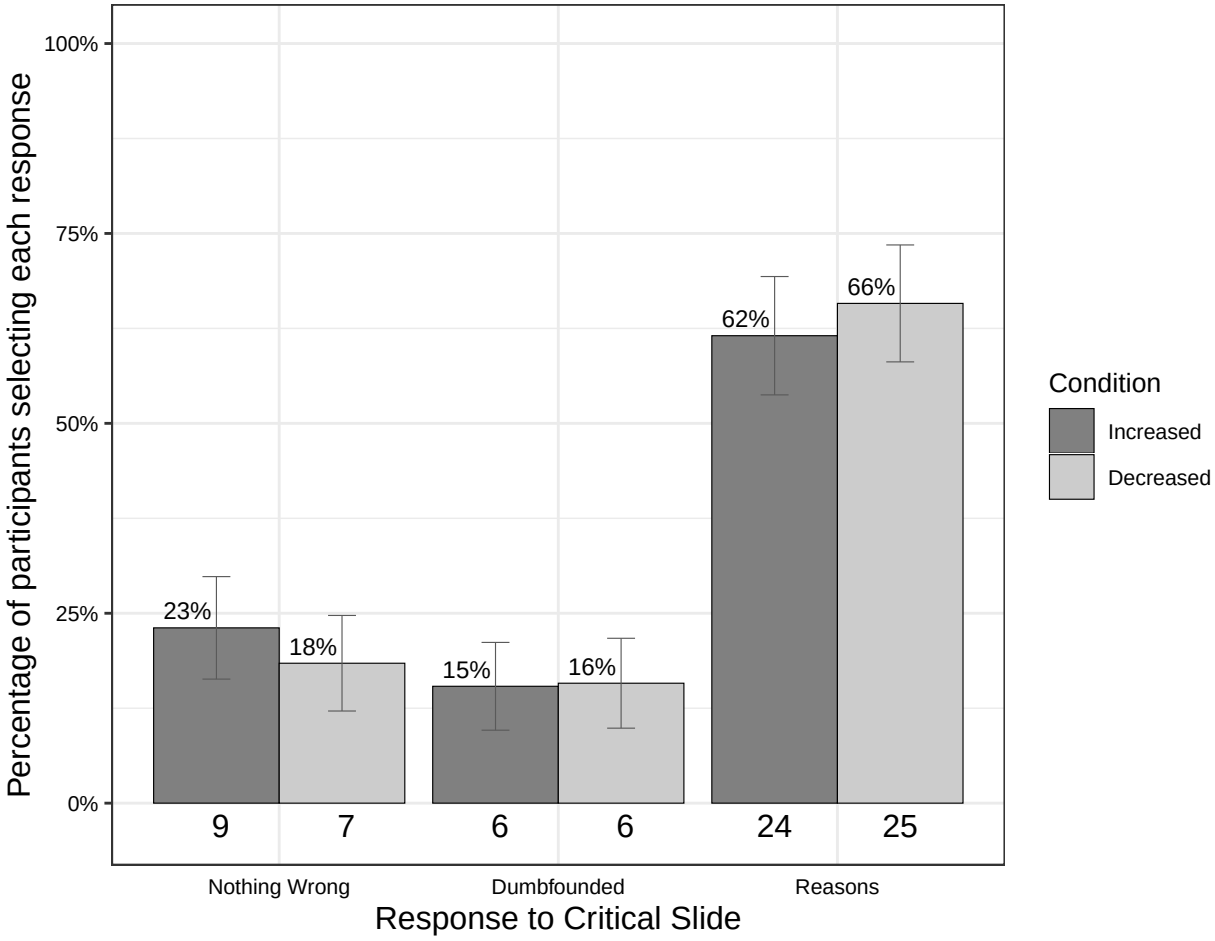
Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(137.88)$ = -0.9, p = .369, d = 0.15, ($M_{\text{increased}}$ = 2.7, $SD_{\text{increased}}$ = 2.2, $M_{\text{decreased}}$ = 3, $SD_{\text{decreased}}$ = 2.1). There was no difference in revised judgement depending on distance manipulation: $t(139.02)$ = -0.52, p = .601, d = 0.09, ($M_{\text{increased}}$ = 2.9, $SD_{\text{increased}}$ = 2.1, $M_{\text{decreased}}$ = 3, $SD_{\text{decreased}}$ = 2.2).

Distancing and Reason-Giving/Dumbfounding (no exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2,\ N = 142) = 0.599$, $p = .741$, $V = 0.06$, the observed power was 0.1. The responses to the critical slide for the experimental group ($N = 67$) and the control group ($N = 75$) are displayed in Figure 8.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4,\ N = 142) = 4.32$, $p = .365$, The observed power was 0.44.

Figure 9

Pilot Study 4: Responses to critical slide for the increased temporal distance group (N = 39) and the reduced temporal distance group (N = 38); (With exclusions; error bars represent standard error of the proportion)



With Exclusions

Overview of Judgments (with exclusions). Fifty seven participants (74.03%) rated the behavior of Julie and Mark as wrong initially, and Fifty eight participants (75.32%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.4, SD = 2.1) and revised ratings (M = 2.4, SD = 2.1), $t(76) = 0.31$, $p = .760$, $d = 0.03$.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(74.98)$ = -0.29, $p = .771$, $d = 0.07$, ($M_{\text{distanced}} = 2.3$, $SD_{\text{distanced}} = 2.1$, $M_{\text{decreased}} = 2.4$, $SD_{\text{decreased}} = 2$). There was no difference in revised judgement depending on distance manipulation: $t(74.95) = 0.14$, $p = .887$, $d = 0.03$, ($M_{\text{increased}} = 2.4$, $SD_{\text{increased}} = 2.1$, $M_{\text{decreased}} = 2.3$, $SD_{\text{decreased}} = 2.1$).

Distancing and Reason-Giving/Dumbfounding (with exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 77) = 0.257$, $p = .879$, $V = 0.06$, the observed power was 0.07. The responses to the critical slide for the experimental group ($N = 39$) and the control group ($N = 38$) are displayed in Figure 9.

Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4, N = 77) = 7.11$, $p = .130$, The observed power was 0.66.

Pilot Study 5 - Temporal Distance, "Julie and Mark" (First Person)

For Pilot Studies 3 and 4, we manipulated temporal distance however participants were asked to think about the scenario from the perspective of a third person. This third person perspective means that increased social distance may have impacted the results. Pilot Study 5 was intended to test the influence of temporal distance on participants' ability to justify their judgments when thinking about them in the first person.

Pilot Study 5: Method

Pilot Study 5: Participants and Design

Pilot Study 5 was a between-subjects design. The dependent variable was rates of providing reasons/dumbfounding (again measured using the critical slide). The independent variable was temporal distance with two levels: increased and decreased. All participants read a vignette similar to the distancing manipulation used in Pilot Studies 3 and 4 that manipulated temporal distance by varying the deadline by which the assignment would need to be completed by - either five weeks in the future (increased) or today (decreased). Unlike previous studies, this was written for participants to think about in the First person. Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

A total sample of 93 participants (53 female, 40 male; $M_{\text{age}} = 38.1$, min = 21, max = 77, $SD = 12.3$) took part. As in Pilot Study 2, participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 71 participants (44 female, 27 male; $M_{\text{age}} = 39.6$, min = 22, max = 65, $SD = 12.1$). Participants were recruited through MTurk in the same way as in Pilot Study 1 (same payment amount, same country selection).

Pilot Study 5: Procedure and Materials

The procedure and materials were similar to Pilot Studies 3 and 4 with a change to the distance manipulation. Rather than a third person perspective (Anne),

participants in Study 5 were asked to consider the "Julie and Mark" vignette in the first person. Temporal distance was manipulated by presenting participants with different imagined deadlines. The experimental manipulations read as follows:

Increased Temporal Distance.

Imagine you taking a philosophy course on ethics. For this ethics course you have been asked to study the scenario below.

You must identify if the behaviour described is right or wrong, and provide reasons for your judgement. **Imagine you must complete and submit this assignment within the next five weeks**.

Decreased Temporal Distance:.

Imagine you taking a philosophy course on ethics. For this ethics course you have been asked to study the scenario below.

You must identify if the behaviour described is right or wrong, and provide reasons for your judgement. **Imagine you must complete and submit this assignment today**.

Pilot Study 5: Results

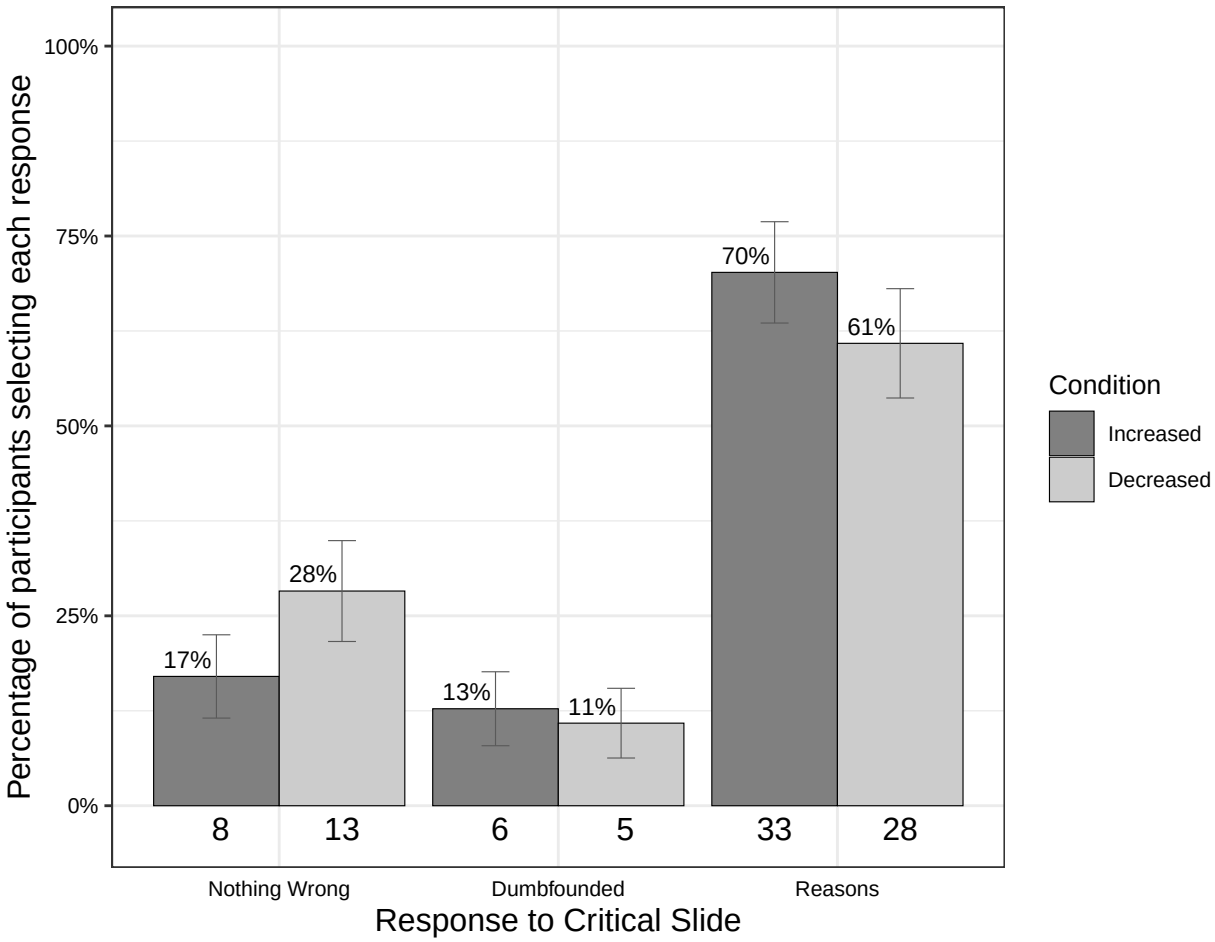
As with previous studies, we present two sets of results. First we present the results for the full sample, second we present the results for the cleaned sample with exclusions based on participants' "lie score".

No Exclusions

Overview of Judgments (no exclusions). Seventy participants (75.27%) rated the behavior of Julie and Mark as wrong initially, and Seventy participants (75.27%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.2, SD = 1.9) and revised ratings (M = 2.3, SD = 1.9), $t(92)$ = -0.47, p = .642, d = 0.05.

Figure 10

Pilot Study 5: Responses to critical slide for the increased temporal distance group (N = 47) and the reduced temporal distance group (N = 46); (No exclusions; error bars represent standard error of the proportion)



Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(83.42) = -1.62$, $p = .109$, $d = 0.34$, ($M_{\text{increased}} = 1.9$, $SD_{\text{increased}} = 1.6$, $M_{\text{decreased}} = 2.5$, $SD_{\text{decreased}} = 2.1$). There was no difference in revised judgement depending on distance manipulation: $t(84.7) = -1.48$, $p = .144$, $d = 0.31$, ($M_{\text{increased}} = 2$, $SD_{\text{increased}} = 1.6$, $M_{\text{decreased}} = 2.5$, $SD_{\text{decreased}} = 2.1$).

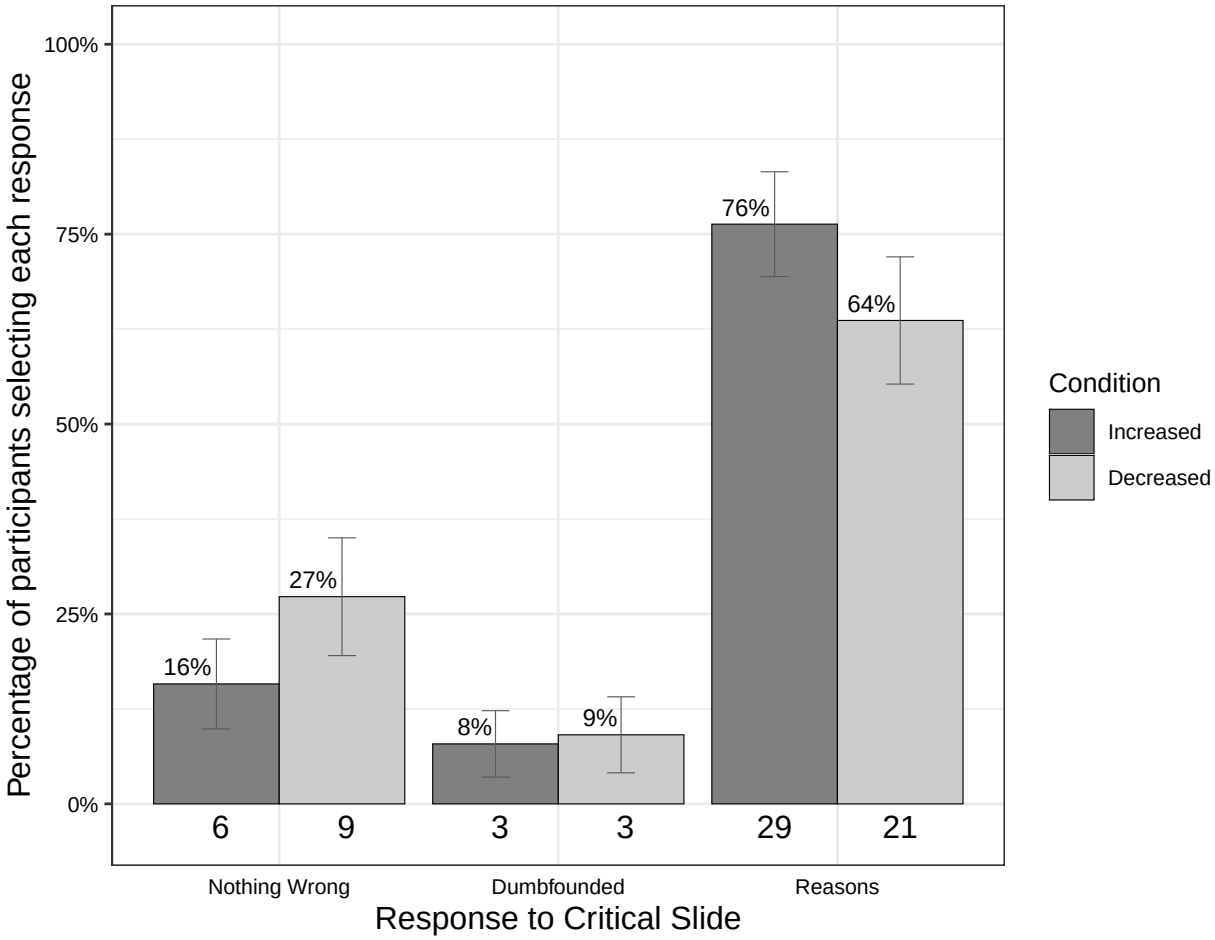
Distancing and Reason-Giving/Dumbfounding (no exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 93) = 1.681$, $p = .432$, $V = 0.13$, the observed power was 0.2. The responses to the critical slide for the experimental group ($N = 47$) and the control

group ($N = 46$) are displayed in Figure 10.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4, N = 93) = 4.92$, $p = .296$, The observed power was 0.5.

Figure 11

Pilot Study 5: Responses to critical slide for the increased temporal distance group (N = 38) and the reduced temporal distance group (N = 33); (With exclusions; error bars represent standard error of the proportion)



With Exclusions

Overview of Judgments (with exclusions). Fifty six participants (78.87%) rated the behavior of Julie and Mark as wrong initially, and Fifty five participants (77.46%) rated the behavior as wrong at the end of the task. There was no significant

difference between initial ratings (M = 2.1, SD = 1.9) and revised ratings (M = 2.2, SD = 1.8), $t(70)$ = -1.76, p = .083, d = 0.21.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(57.14)$ = -1.41, p = .164, d = 0.34, ($M_{\text{distanced}}$ = 1.8, $SD_{\text{distanced}}$ = 1.5, $M_{\text{decreased}}$ = 2.4, $SD_{\text{decreased}}$ = 2.1). There was no difference in revised judgement depending on distance manipulation: $t(58.89)$ = -1.67, p = .100, d = 0.41, ($M_{\text{increased}}$ = 1.8, $SD_{\text{increased}}$ = 1.6, $M_{\text{decreased}}$ = 2.6, $SD_{\text{decreased}}$ = 2.1).

Distancing and Reason-Giving/Dumbfounding (with exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 71) = 1.536$, p = .464, V = 0.15, the observed power was 0.18. The responses to the critical slide for the experimental group (N = 38) and the control group (N = 33) are displayed in Figure 7.

Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4, N = 71) = 5.15$, p = .272, The observed power was 0.52.

Pilot Study 6 - Temporal Distance, "Jennifer" (Third Person)

Pilot Study 6 was similar to the previous Pilot Studies, with two notable changes. First, across Pilot Studies 1-5 participants responded to the most widely used dumbfounding scenario, the *classic* "Julie and Mark" scenario (McHugh et al., 2023). In Pilot Study 6 we replaced the "Julie and Mark" scenario with the "Jennifer" scenario, which describes an act of cannibalism involving a cadaver that is already dead. Second, in all previous Pilot Studies dumbfounding was measured using the the critical slide which contained three categorical response options. The categorical nature of the response options means that this measure may not be well suited to detecting small effects. To address this possibility we trialed a scale measure of dumbfounding in Pilot Study 6.

We manipulated temporal distance using the same "Anne" vignette as in Pilot Studies 3 and 4. The aim of Pilot Study 6 was to investigate if temporal distance influenced participants ability to justify their judgments in the "Jennifer" scenario.

Pilot Study 6: Method

Pilot Study 6: Participants and Design

Pilot Study 6 was a between-subjects design. The dependent variable was rates of providing reasons/dumbfounding, measured using a revised critical slide. The independent variable was temporal distance with two levels: increased and decreased. All participants read a brief vignette where temporal distance was manipulated by varying the deadline by which an assignment would need to be completed by - either five weeks in the future (increased) or today (decreased). Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

A total sample of 134 participants (70 female, 64 male; $M_{\text{age}} = 36.3$, min = 21, max = 73, $SD = 12.2$) took part. Again, participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 76 participants (39 female, 37 male; $M_{\text{age}} = 38.1$, min = 21, max = 73, $SD = 12.3$). Participants were

recruited through MTurk under the same conditions as previous Pilot Studies.

Pilot Study 6: Procedure and Materials

The procedure and materials remained similar to previous Pilot Studies, with a change in the scenario being considered and a change to the way dumbfounding was measured.

Participants were presented with the "Jennifer" scenario (below).

Temporal distance was manipulated using the same instructions as in Pilot Studies 3 and 4.

Jennifer Scenario.

> Jennifer works in a medical school pathology lab as a research assistant. The lab prepares human cadavers that are used to teach medical students about anatomy. The cadavers come from people who had donated their body to science for research. One night Jennifer is leaving the lab when she sees a body that is going to be discarded the next day. Jennifer was a vegetarian, for moral reasons. She thought it was wrong to kill animals for food. But then, when she saw a body about to be cremated, she thought it was irrational to waste perfectly edible meat. So she cut off a piece of flesh, and took it home and cooked it. The person had died recently of a heart attack, and she cooked the meat thoroughly, so there was no risk of disease

Increased Temporal Distance.

> Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.
>
> She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her

judgement. **She must complete and submit this assignment within the next five weeks**.

While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Decreased Temporal Distance:.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her judgement. **She must complete and submit this assignment today**.

While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Dumbfounding was measured using a revised version of the *critical slide*, which presented a statement defending the actions of Jennifer followed by two questions, one regarding participants judgments, and one regarding participants ability to justify their response, responses to both questions were recorded on 7 point Likert scales. The revised slide read as follows: "Jennifer's actions did not harm anyone, or negatively affect anyone. How can there be anything wrong with what she did?". Participants responded on a 7-point Likert scale (1 = *There is nothing wrong*; 7 = *It is definitely wrong*). Below this, the following question was presented: "Please indicate how well you can justify your response above", again, participants responded on a 7-point Likert scale (1 = *I can provide a valid reason*; 7 = *I can't think of a reason*).

Pilot Study 6: Results

Again, we present two sets of results. First we present the results for the full sample, second we present the results for the cleaned sample with exclusions based on participants' "lie score".

No Exclusions

Overview of Judgments (no exclusions). Ninety two participants (68.66%) rated the behavior of Jennifer as wrong initially, and Eighty nine participants (66.42%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.7, SD = 2.1) and revised ratings (M = 2.7, SD = 2.1), $t(133)$ = -0.84, p = .401, d = 0.07.

Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(131.9)$ = -0.34, p = .736, d = 0.06, ($M_{\text{increased}}$ = 2.6, $SD_{\text{increased}}$ = 2.1, $M_{\text{decreased}}$ = 2.7, $SD_{\text{decreased}}$ = 2.2). There was no difference in revised judgement depending on distance manipulation: $t(131.7)$ = -0.62, p = .538, d = 0.11, ($M_{\text{increased}}$ = 2.6, $SD_{\text{increased}}$ = 2, $M_{\text{decreased}}$ = 2.8, $SD_{\text{decreased}}$ = 2.2).

Distancing and Reason-Giving/Dumbfounding (no exclusions). At the point when dumbfounding was recorded, there was no significant difference in judgment depending on the distance manipulation, $t(126.08)$ = 0.91, p = .367 ($M_{\text{increased}}$ = 5.9, $SD_{\text{increased}}$ = 1.5, $M_{\text{decreased}}$ = 5.6, $SD_{\text{decreased}}$ = 1.9). There was also no difference self report ability to provide reasons depending on distance manipulation, $t(131.33)$ = 1.03, p = .306 ($M_{\text{increased}}$ = 3.3, $SD_{\text{increased}}$ = 2.1, $M_{\text{decreased}}$ = 2.9, $SD_{\text{decreased}}$ = 2.1).

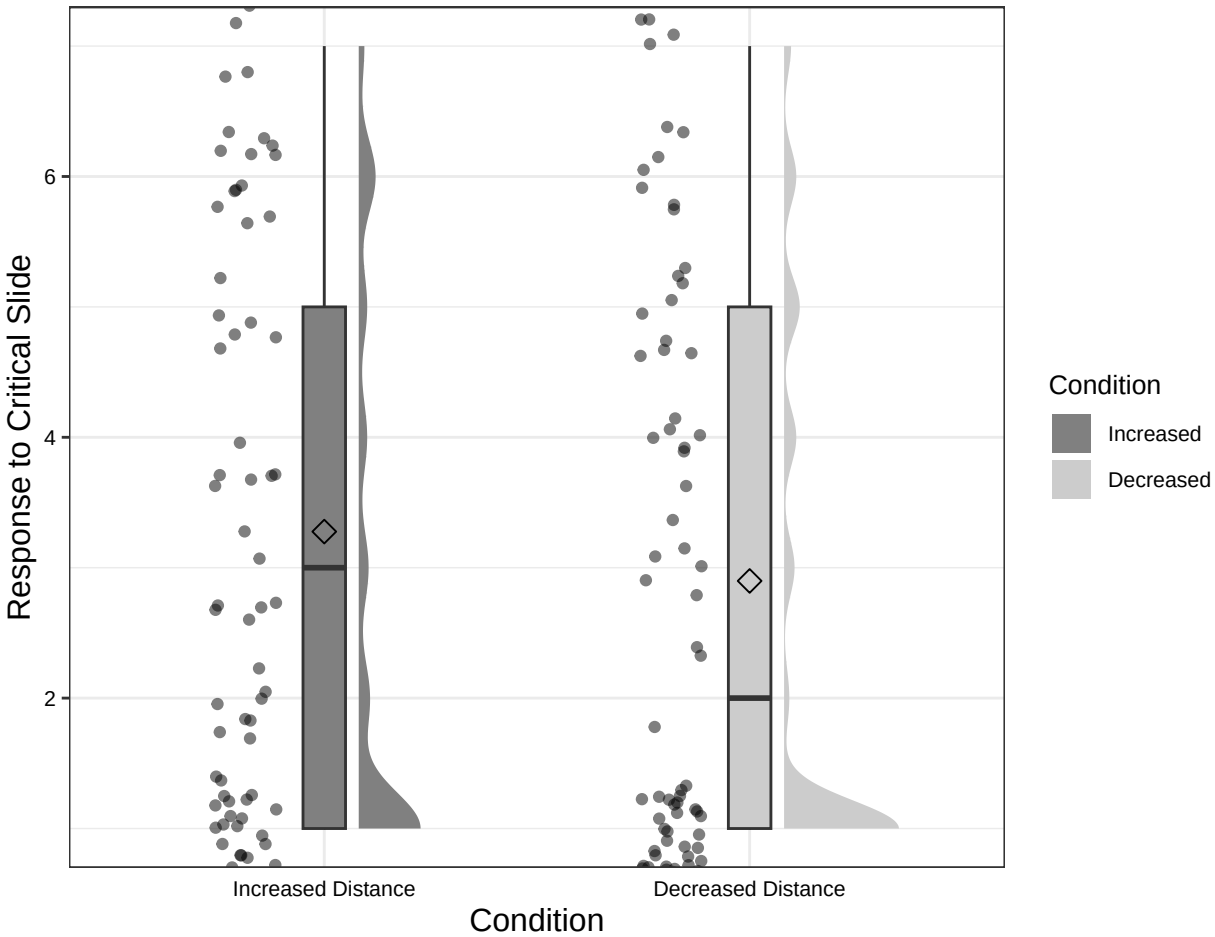
Individual Difference Predictors (no exclusions). Two linear regressions were conducted to investigate the potential influence of need for closure and cognitive reflection on both responses on the revised critical slide.

For the first regression the outcome variable was judgment. Overall the model did not significantly predict participants judgments on the revised critical slide, $R^2 =< .01$, $F(2, 131) = 0.61$, $p = .544$.

For the second regression the outcome variable was self-report ability to provide

Figure 12

*Pilot Study 6: Responses to critical slide for the increased temporal distance group (N = 65) and the reduced temporal distance group (N = 69); (No exclusions; 1 = *I can provide a valid reason*; 7 = *I can't think of a reason*)*



reasons. Overall the model significantly predicted participants self-reported ability to provide reasons, $R^2 = .13$, $F(2, 131) = 9.68$, $p < .001$. Need for closure was a significant predictor in the model, $b = -0.03$, 95% CI $[-0.05, -0.02]$, $t(131) = -3.78$, $p < .001$, while cognitive reflection was not $b = -0.20$, 95% CI $[-0.51, 0.10]$, $t(131) = -1.30$, $p = .196$.

With Exclusions

Overview of Judgments (with exclusions). Sixty six participants (86.84%) rated the behavior of Jennifer as wrong initially, and Sixty three participants (82.89%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 1.9, SD = 1.7) and revised ratings (M = 2.1, SD = 1.9), $t(75)$ = -1.89, p = .063, d = 0.22.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(71.25)$ = -1.75, p = .084, d = 0.38, ($M_{\text{increased}}$ = 1.5, $SD_{\text{increased}}$ = 1.2, $M_{\text{decreased}}$ = 2.2, $SD_{\text{decreased}}$ = 2). There was no difference in revised judgement depending on distance manipulation: $t(73.95)$ = -1.13, p = .263, d = 0.25, ($M_{\text{increased}}$ = 1.8, $SD_{\text{increased}}$ = 1.5, $M_{\text{decreased}}$ = 2.2, $SD_{\text{decreased}}$ = 2.1).

Distancing and Reason-Giving/Dumbfounding (with exclusions). At the point when dumbfounding was recorded, there was no significant difference in judgment depending on the distance manipulation, $t(73.86)$ = 0.93, p = .355 ($M_{\text{increased}}$ = 6.2, $SD_{\text{increased}}$ = 1.5, $M_{\text{decreased}}$ = 5.8, $SD_{\text{decreased}}$ = 2). There was also no difference self report ability to provide reasons depending on distance manipulation, $t(63.14)$ = 1.09, p = .278 ($M_{\text{increased}}$ = 2.7, $SD_{\text{increased}}$ = 2.1, $M_{\text{decreased}}$ = 2.2, $SD_{\text{decreased}}$ = 1.9).

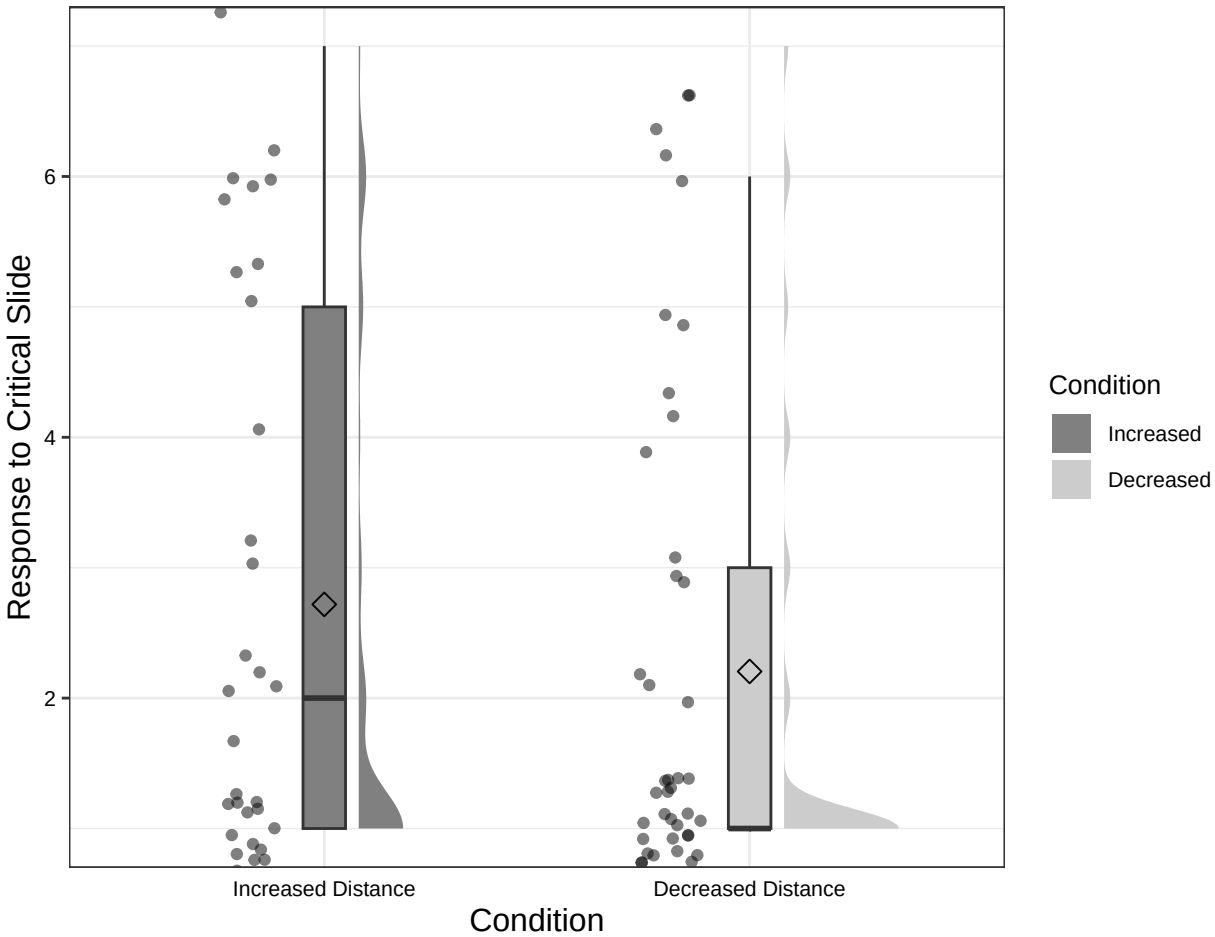
Individual Difference Predictors (with exclusions). Two linear regressions were conducted to investigate the potential influence of need for closure and cognitive reflection on both responses on the revised critical slide.

For the first regression the outcome variable was judgment. Overall the model did not significantly predict participants judgments on the revised critical slide, $R^2 =< .01$, $F(2, 73) = 0.08$, $p = .924$.

For the second regression the outcome variable was self-report ability to provide reasons. Overall the model significantly predicted participants self-reported ability to provide reasons, $R^2 = .08$, $F(2, 73) = 3.38$, $p = .040$. Need for closure was a significant predictor in the model, $b = -0.03$, 95% CI $[-0.05, -0.01]$, $t(73) = -2.57$, $p = .012$, while cognitive reflection was not $b = 0.14$, 95% CI $[-0.24, 0.51]$, $t(73) = 0.73$, $p = .470$.

Figure 13

*Pilot Study 6: Responses to critical slide for the increased temporal distance group (N = 32) and the reduced temporal distance group (N = 44); (With exclusions; 1 = *I can provide a valid reason*; 7 = *I can't think of a reason*)*



Pilot Study 7 - Temporal Distance, "Jennifer" (Third Person)

Pilot Study 6 did not provide a strong case for adopting the continuous measure of dumbfounding in the revised critical slide. As such, Pilot Study 7 reverted to the validated critical slide (McHugh et al., 2017, 2020) to measure dumbfounded responding. We used the "Jennifer" scenario again. The aim of Study 7 was to test the influence of temporal distancing on participants' ability to justify their judgments of the "Jennifer" scenario.

Pilot Study 7: Method

Pilot Study 7: Participants and Design

Pilot Study 7 was a between-subjects design. The dependent variable was rates of reason-giving/dumbfounding measured using the critical slide (with three response options, reason-giving, dumbfounding, and nothing-wrong). The independent variable was temporal distance with two levels: increased and decreased. All participants the distancing manipulation vignette used in Pilot Studies 3 and 4. Temporal distance was manipulated by varying the deadline by which the assignment would need to be completed by - either five weeks in the future (increased) or today (decreased). Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

A total sample of 138 participants (87 female, 51 male; $M_{\text{age}} = 36.8$, min = 20, max = 68, $SD = 12$) took part. Again, participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 87 participants (64 female, 23 male; $M_{\text{age}} = 39.2$, min = 20, max = 67, $SD = 12$). Participants were recruited through MTurk under the same conditions as in previous pilot studies.

Pilot Study 7: Procedure and Materials

The procedure and materials were similar to Pilot Studies 3 and 4 with a single change. Rather than the "Julie and Mark" scenario, participants were asked to consider the "Jennifer" dilemma used in Pilot Study 6. The scenario and distance manipulations read as follows.

Jennifer Scenario.

Jennifer works in a medical school pathology lab as a research assistant. The lab prepares human cadavers that are used to teach medical students about anatomy. The cadavers come from people who had donated their body to science for research. One night Jennifer is leaving the lab when she sees a body that is going to be discarded the next day. Jennifer was a vegetarian, for moral reasons. She thought it was wrong to kill animals for food. But then, when she saw a body about to be cremated, she thought it was irrational to waste perfectly edible meat. So she cut off a piece of flesh, and took it home and cooked it. The person had died recently of a heart attack, and she cooked the meat thoroughly, so there was no risk of disease

Increased Temporal Distance.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her judgement. **She must complete and submit this assignment within the next five weeks**.

While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Decreased Temporal Distance:.

Anne is a student of philosophy. She generally shows a good understanding of the subject matter, and this is reflected in her grades. Sometimes, however, she may adopt a position on an issue and struggle (or even fail) to defend it.

She is currently taking a course in ethics. For this ethics course, Anne and has been asked to study the following scenario. She must identify if the behaviour described is right or wrong, and provide reasons for her judgement. **She must complete and submit this assignment today**.

While reading the story on the next page, try to think about the story from Anne's perspective rather than your own.

Pilot Study 7: Results

Again, we present two sets of results below. First we present the results for the full sample, second we present the results for the cleaned sample with exclusions based on participants' "lie score".

No Exclusions

Overview of Judgments (no exclusions). Ninety seven participants (70.29%) rated the behavior of Julie and Mark as wrong initially, and Ninety three participants (67.39%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.4, SD = 2) and revised ratings (M = 2.5, SD = 2), $t(137)$ = -0.72, p = .470, d = 0.06.

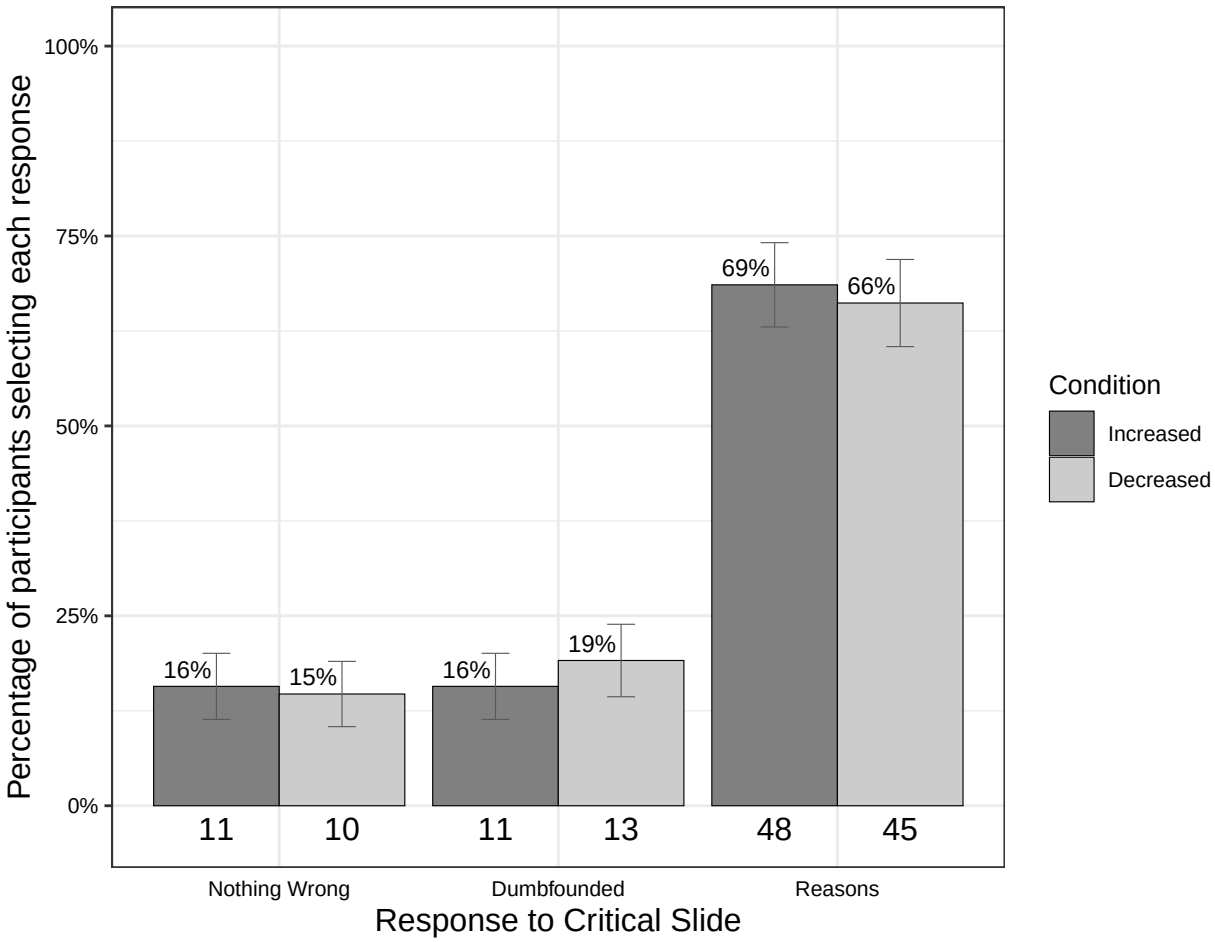
Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(136)$ = 0.6, p = .547, d = 0.1, ($M_{\text{increased}}$ = 2.5, $SD_{\text{increased}}$ = 2, $M_{\text{decreased}}$ = 2.3, $SD_{\text{decreased}}$ = 2). There was no difference in revised judgement depending on distance manipulation: $t(135.79)$ = -0.04, p = .966, d = 0.01, ($M_{\text{increased}}$ = 2.5, $SD_{\text{increased}}$ = 2, $M_{\text{decreased}}$ = 2.5, $SD_{\text{decreased}}$ = 2).

Distancing and Reason-Giving/Dumbfounding (no exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 138) = 0.282$, p = .868, V = 0.05, the observed power was 0.07. The responses to the critical slide for the experimental group (N = 70) and the control group (N = 68) are displayed in Figure 14.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted

Figure 14

Pilot Study 7: Responses to critical slide for the increased temporal distance group (N = 70) and the reduced temporal distance group (N = 68); (No exclusions; error bars represent standard error of the proportion)



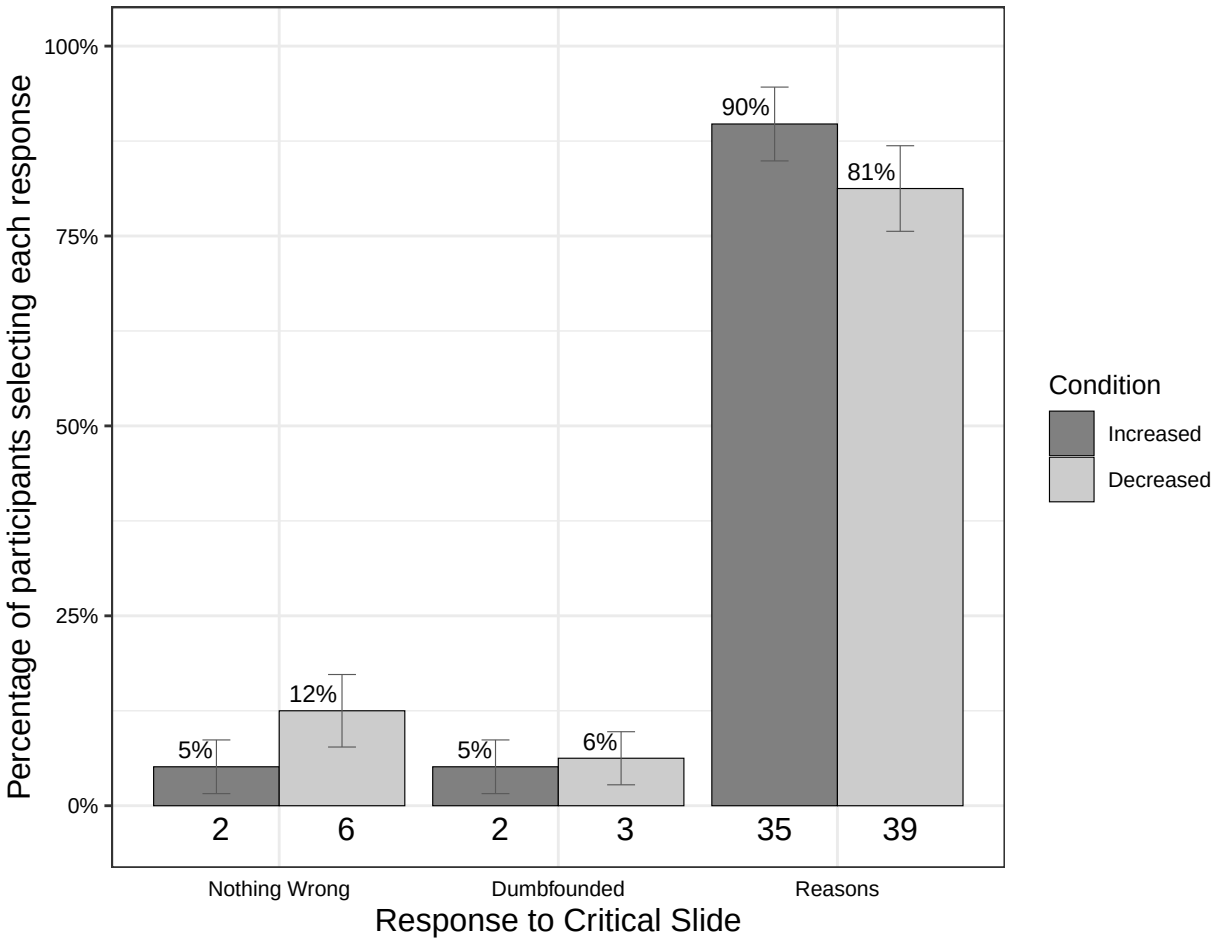
dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4,\ N = 138) = 7.74$, $p = .102$, The observed power was 0.7.

With Exclusions

Overview of Judgments (with exclusions). Seventy nine participants (90.8%) rated the behavior of Julie and Mark as wrong initially, and Seventy seven participants (88.51%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings ($M = 1.6$, $SD = 1.3$) and revised ratings ($M = 1.6$, $SD = 1.3$), $t(86) = -1.09$, $p = .278$, $d = 0.12$.

Figure 15

Pilot Study 7: Responses to critical slide for the increased temporal distance group (N = 39) and the reduced temporal distance group (N = 48); (With exclusions; error bars represent standard error of the proportion)



Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(79.44) = -0.79$, $p = .433$, $d = 0.16$, ($M_{\text{distanced}} = 1.5$, $SD_{\text{distanced}} = 0.9$, $M_{\text{decreased}} = 1.7$, $SD_{\text{decreased}} = 1.5$). There was no difference in revised judgement depending on distance manipulation: $t(80.22) = -1.14$, $p = .259$, $d = 0.23$, ($M_{\text{increased}} = 1.5$, $SD_{\text{increased}} = 1$, $M_{\text{decreased}} = 1.8$, $SD_{\text{decreased}} = 1.5$).

Distancing and Reason-Giving/Dumbfounding (with exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 87) = 1.501$, $p = .472$, $V = 0.13$, the observed power was

0.18. The responses to the critical slide for the experimental group ($N = 39$) and the control group ($N = 48$) are displayed in Figure 15.

Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model did not significantly predict responses to the critical slide $\chi^2(4, N = 87) = 3.52$, $p = .475$, The observed power was 0.37.

Pilot Study 8 - Temporal Distance, "Julie and Mark" (First Person)

All previous manipulations of temporal distance involved an additional vignette that included instructions to consider the moral scenario on either an extended timeline (increased temporal distance), or a short timeline (reduced temporal distance). In Pilot Study 8 we embedded the distance manipulation into the main moral scenario being judged. The scenario was depicted as either happening in the distant past (increased temporal distance) or no information was provided about when the event happened (control). We hypothesized that for the increased temporal distance condition rates of reason-giving would be higher.

Pilot Study 8: Method

Pilot Study 8: Participants and Design

Pilot Study 8 was a between-subjects design. The dependent variable was rates of providing reasons/dumbfounding measured using the critical slide (with three response options, reason-giving, dumbfounding, and nothing-wrong). The independent variable was temporal distance with two levels: increased and control. All participants read a modified version of the "Julie and Mark" vignette. Temporal distance was manipulated by the addition of a sentence indicating the event happened in the distant past (60 years ago). The control condition did not include information about when the event happened. Need for closure (Kruglanski, 1996, 2013; Kruglanski et al., 2013) and cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak et al., 2011) were recorded as potential correlate/moderator variables.

An initial sample of 286 participants took part. After initial screening (3 attention checks) we were left with a sample of 244 participants (135 female, 109 male; $M_{\text{age}} = 37.9$, min = 21, max = 69, $SD = 12.4$) took part. When participants who scored above the threshold on the need for closure lie score were removed, and this left a sample of 124 participants (69 female, 55 male; $M_{\text{age}} = 42.3$, min = 21, max = 68, $SD = 12.7$). Participants were recruited through MTurk under the same conditions as in previous pilot studies.

Pilot Study 8: Procedure and Materials

The procedure was similar to previous studies with some changes. The distance manipulation was embedded into the moral scenario rather than through an additional vignette. Reason-giving/dumbfounded responding was recorded using the critical slide, with three response options (reason-giving, nothing-wrong, dumbfounded). All participants read an rated the "Julie and Mark" scenario. In the experimental (increased temporal distance) condition this was modified to include information about when the event took place. The control condition did not include this information. The scenarios read as follows:

Increased Temporal Distance.

60 years ago, Julie and Mark, who are brother and sister, were travelling together in France. They were both on summer vacation from college. One night they were staying alone in a cabin near the beach. They decided that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark used a condom too, just to be safe. They both enjoyed it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other.

Control.

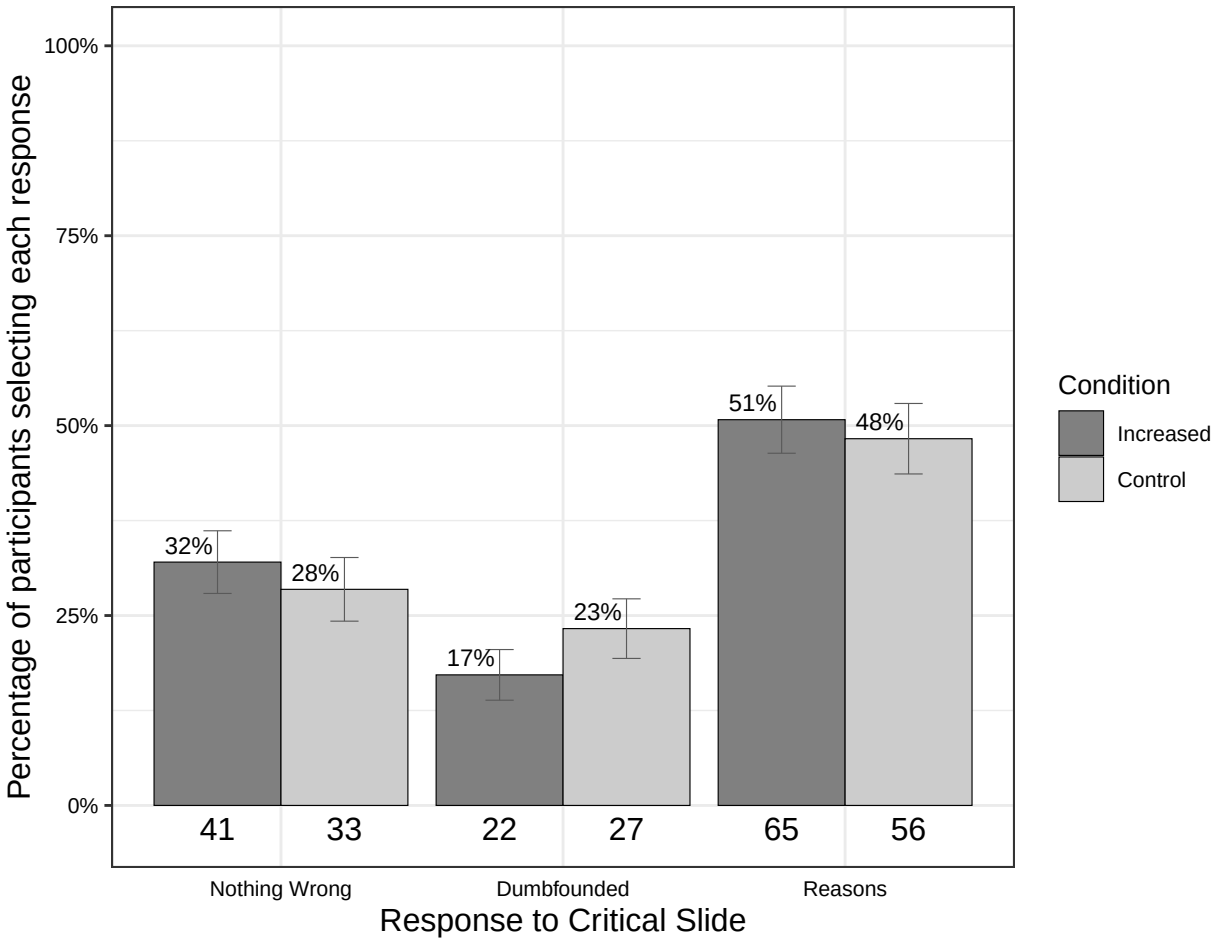
Julie and Mark, who are brother and sister, are travelling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other.

Pilot Study 8: Results

Again, we present two sets of results below. First we present the results for the full sample, second we present the results for the cleaned sample with exclusions based on participants' "lie score".

Figure 16

Pilot Study 8: Responses to critical slide for the increased temporal distance group (N = 128) and the control group (N = 116); (No exclusions; error bars represent standard error of the proportion)



No Exclusions

Overview of Judgments (no exclusions). one hundred Forty nine participants (61.07%) rated the behavior of Julie and Mark as wrong initially, and one hundred Forty six participants (59.84%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 3, SD = 2.2) and

revised ratings (M = 2.9, SD = 2.1), $t(243)$ = 0.84, p = .403, d = 0.05.

Distancing and Judgments Made (no exclusions). There was no difference in initial judgement depending on distance manipulation: $t(235.34)$ = -0.34, p = .733, d = 0.04, ($M_{\text{increased}}$ = 2.9, $SD_{\text{increased}}$ = 2.1, M_{control} = 3, SD_{control} = 2.3). There was no difference in revised judgement depending on distance manipulation: $t(240.71)$ = 0.9, p = .366, d = 0.12, ($M_{\text{increased}}$ = 3.1, $SD_{\text{increased}}$ = 2.1, M_{control} = 2.8, SD_{control} = 2.1).

Distancing and Reason-Giving/Dumbfounding (no exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2, N = 244) = 1.458$, p = .482, V = 0.08, the observed power was 0.17. The responses to the critical slide for the experimental group (N = 128) and the control group (N = 116) are displayed in Figure 16.

Individual Difference Predictors (no exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model significantly predicted responses to the critical slide $\chi^2(4, N = 244) = 13.43$, p = .009, The observed power was 0.92. The model explained between 3.68% (Cox and Snell R square) and 5.2% (Nadelkerke R squared) of the variance in responses to the critical slide. Need for closure was a significant predictor in the model. As need for closure increased, participants were significantly more likely to provide reasons than to select nothing-wrong, Wald = 10.48, p = .001, odds ratio = 0.97, 95% CI [0.96, 0.99]. The predicted probabilities of each response depending on need for cognition are displayed in Figure 17.

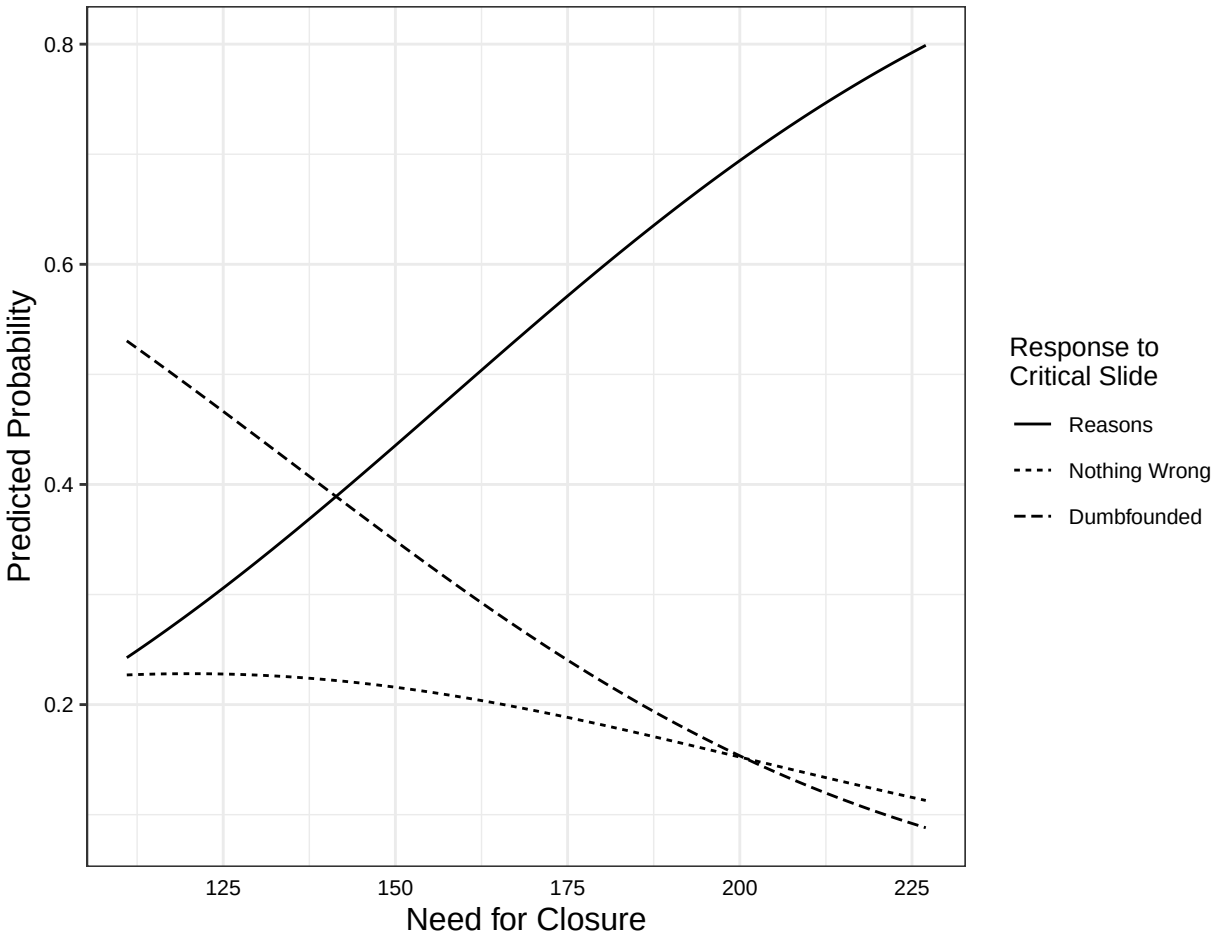
With Exclusions

Overview of Judgments (with exclusions). Ninety two participants (74.19%) rated the behavior of Julie and Mark as wrong initially, and Ninety participants (72.58%) rated the behavior as wrong at the end of the task. There was no significant difference between initial ratings (M = 2.4, SD = 2) and revised ratings (M = 2.4, SD = 2), $t(123)$ = -0.21, p = .832, d = 0.02.

Distancing and Judgments Made (with exclusions). There was no difference in initial judgement depending on distance manipulation: $t(115.16)$ = 0.74, p

Figure 17

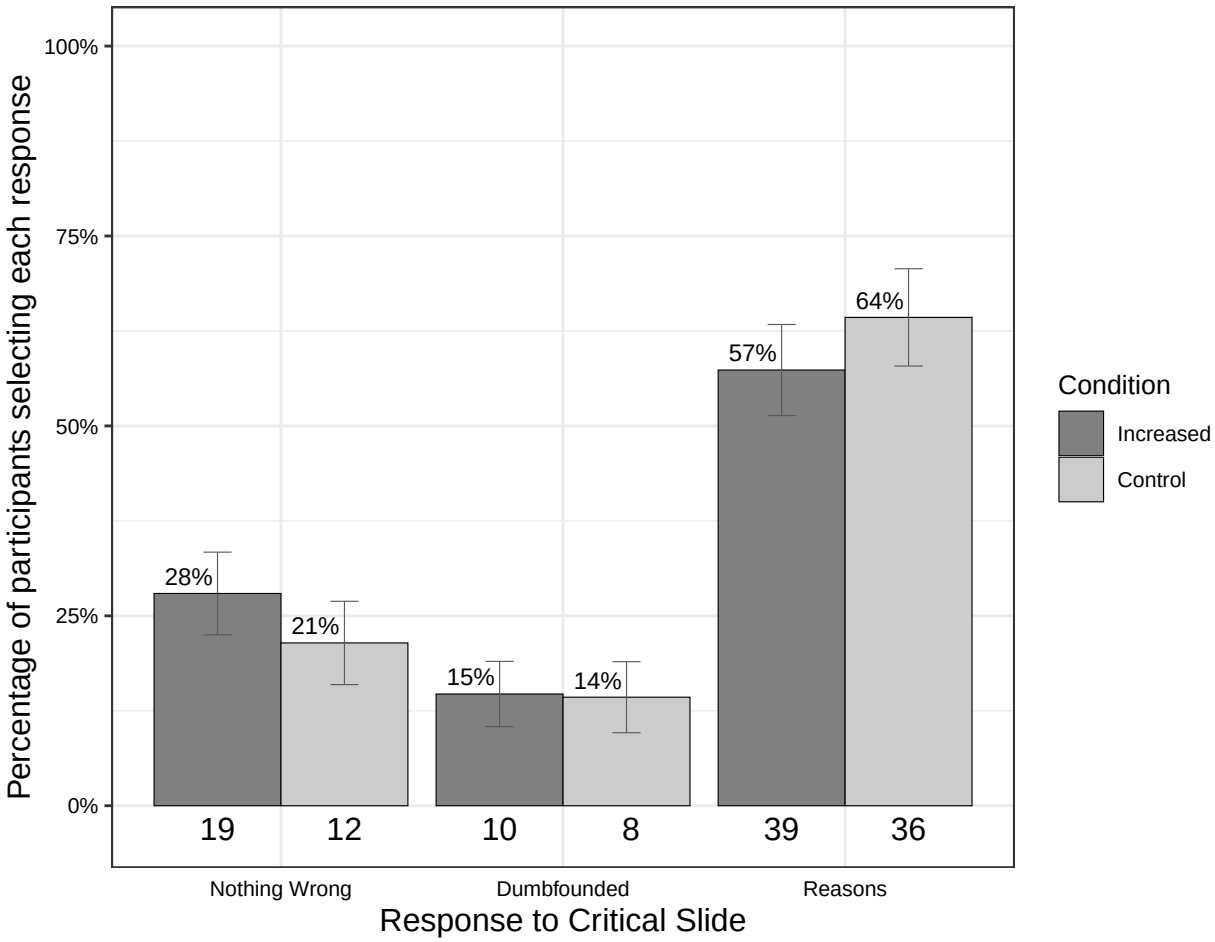
Pilot Study 8: Probability of selecting each response to the critical slide depending on Social Desirability (no exclusions)



$= .458$, $d = 0.13$, ($M_{\text{distanced}} = 2.5$, $SD_{\text{distanced}} = 2$, $M_{\text{decreased}} = 2.3$, $SD_{\text{decreased}} = 2.1$). Revised judgments were significantly more severe in the increased temporal distance condition ($M_{\text{increased}} = 2.7$, $SD_{\text{increased}} = 2.1$) than in the reduced temporal distance condition ($M_{\text{decreased}} = 2.2$, $SD_{\text{decreased}} = 1.8$), $t(121.61) = 1.42$, $p = .158$, $d = 0.25$,

Figure 18

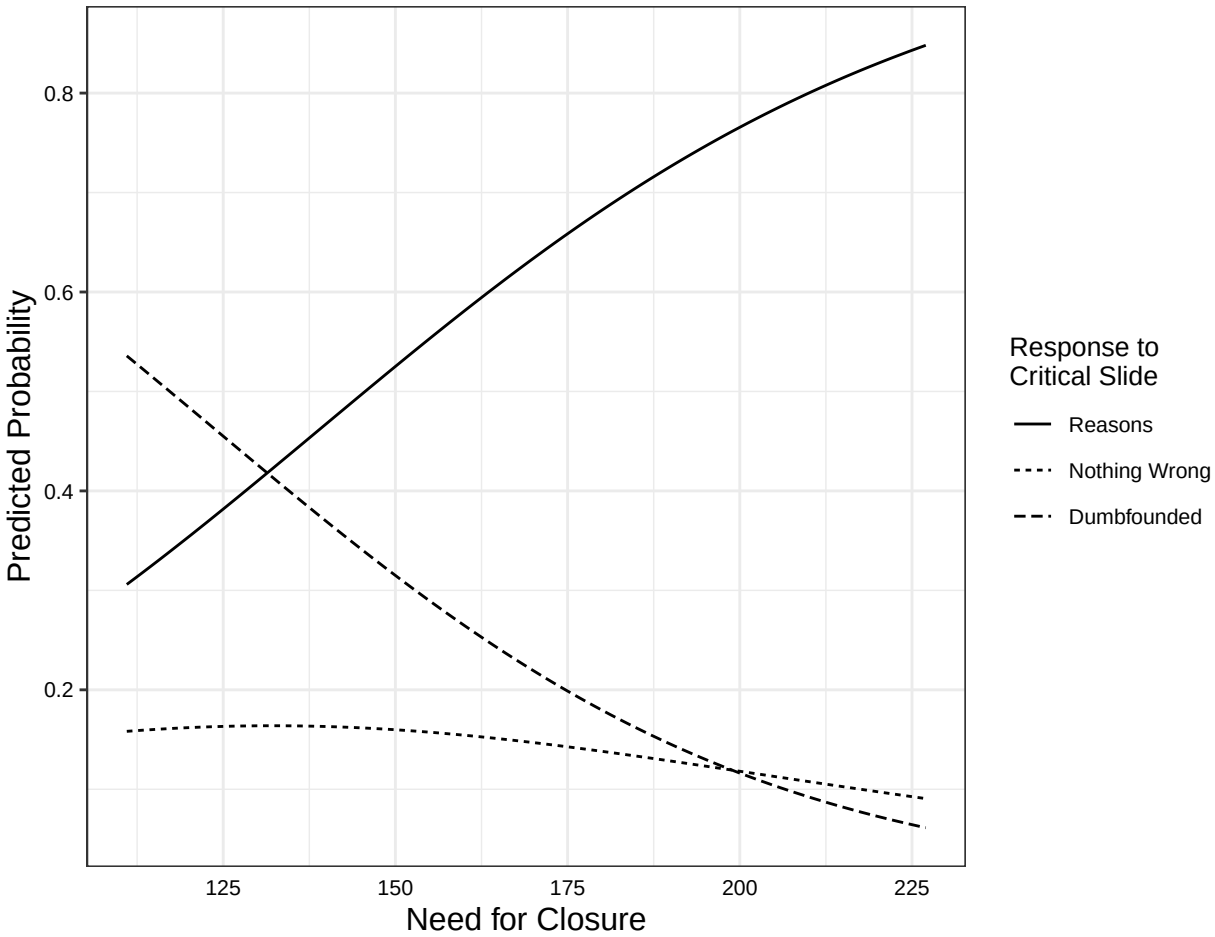
Pilot Study 8: Responses to critical slide for the increased temporal distance group (N = 68) and the control group (N = 56); (With exclusions; error bars represent standard error of the proportion)



Distancing and Reason-Giving/Dumbfounding (with exclusions). There was no significant association between experimental condition and response to the critical slide, $\chi^2(2,\ N = 124) = 0.769$, $p = .681$, $V = 0.08$, the observed power was 0.11. The responses to the critical slide for the experimental group ($N = 68$) and the control group ($N = 56$) are displayed in Figure 18.

Figure 19

Pilot Study 8: Probability of selecting each response to the critical slide depending on need for closure (with exclusions)



Individual Difference Predictors (with exclusions). A multinomial logistic regression was conducted to test if the individual difference measures predicted dumbfounding. Overall the model significantly predicted responses to the critical slide $\chi^2(4, N = 124) = 10.57$, $p = .032$, The observed power was 0.84. The model explained between 3.68% (Cox and Snell R square) and 5.2% (Nadelkerke R squared) of the variance in responses to the critical slide. Need for closure was a significant predictor in the model. As need for closure increased, participants were significantly more likely to provide reasons than to select nothing-wrong, Wald = 10.48, $p = .005$, odds ratio = 0.97, 95% CI [0.96, 0.99]. The predicted probabilities of each response depending on need for cognition are displayed in Figure 19.

Combined Results

Participants

When not excluding participants based on their need for cognition "lie" scores, the combined sample included a total of 918 participants (488 female, 429 male; M_{age} = 37.55, min = 19, max = 83, SD = 12.17)

When participants who exceeded the threshold on the "lie" score were excluded the combined sample included a total of 626 participants (353 female, 273 male; M_{age} = 39.76, min = 19, max = 83, SD = 12.52)

Combined Results: No Exclusions

Combined Increased Distance vs Reduced Distance (no exclusions)

Our first analysis did not distinguish between temporal and social distance. We created a new variable with three levels, increased distance, decreased distance, and control. A chi-squared test for association revealed a significant a significant association between experimental condition and response to the critical slide: $\chi^2(4,\ N = 918) = 11.459$, $p = .022$, $V = 0.11$, the observed power was 0.87 The responses to the critical slide for the control group ($N = 97$), the increased distance group ($N = 464$) and the decreased distance group ($N = 357$) are displayed in Figure 20. The observed counts, expected counts and standardized residuals are displayed in Table 3.

Figure 20

Studies 1, 2, 3, 4, 5, 7, and 8: Responses to critical slide depending on condition

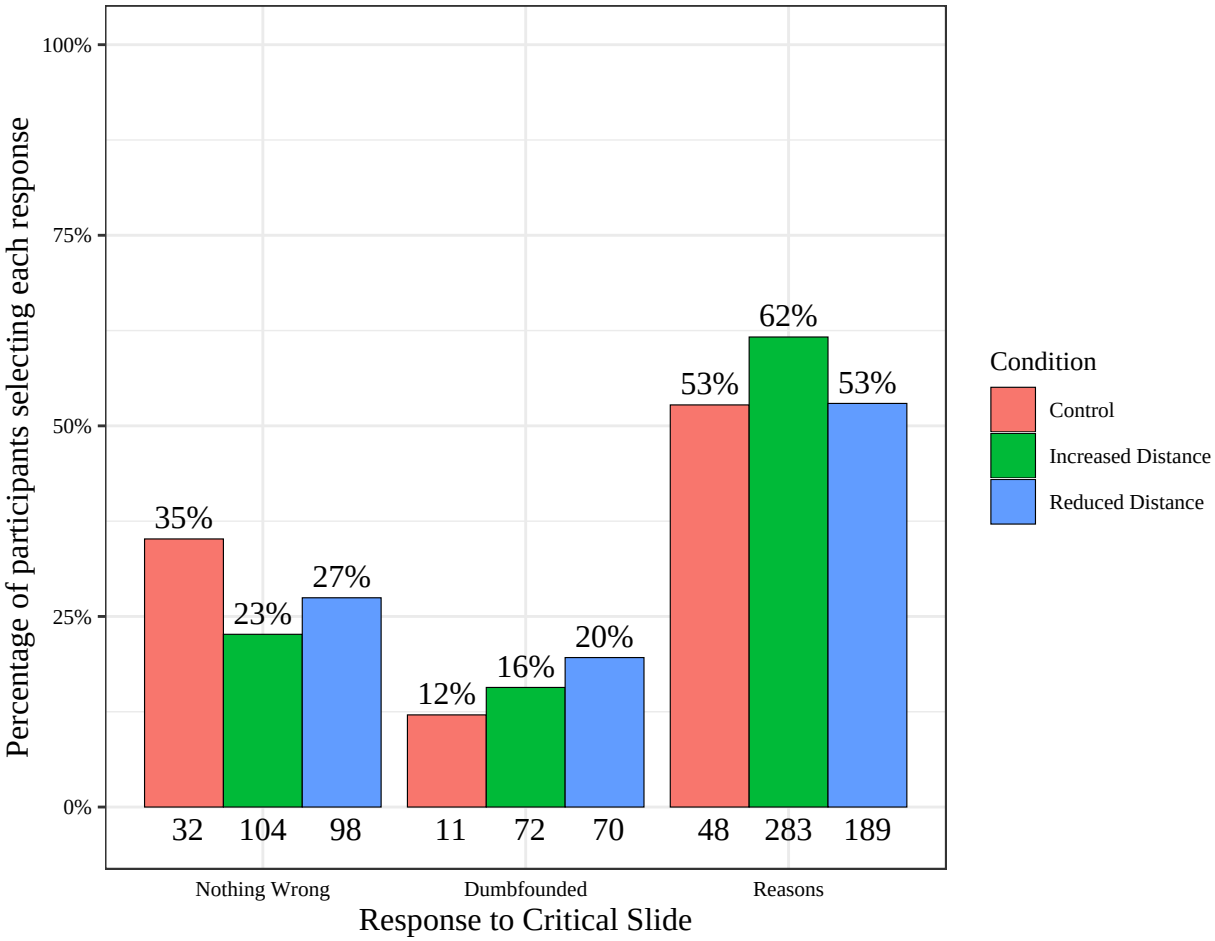


Table 3

Studies 1, 2, 3, 4, 5, 7, and 8 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending distance manipulation

		Control	Long	Short
Observed count	Nothing Wrong	32	104	98
	Dumbfounded	11	72	70
	Reasons	48	283	189
Expected count	Nothing Wrong	23.48	118.42	92.1
	Dumbfounded	15.35	77.43	60.22
	Reasons	52.17	263.15	204.67
Standardised residuals	Nothing Wrong	2.15*	-2.19*	0.92
	Dumbfounded	-1.28	-0.96	1.77
	Reasons	-0.93	2.67*	-2.15*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Different types of manipulations (Temporal / Social)

Overall, there was no significant association between experimental condition and response to the critical slide: $\chi^2(10, N = 907) = 16.372$, $p = .089$, $V = 0.13$, the observed power was 0.96. The responses to the critical slide for each type of manipulation (the total samples are displayed below) are displayed in Figure 21. The observed counts, expected counts and standardized residuals are displayed in Table 4.

Table 4

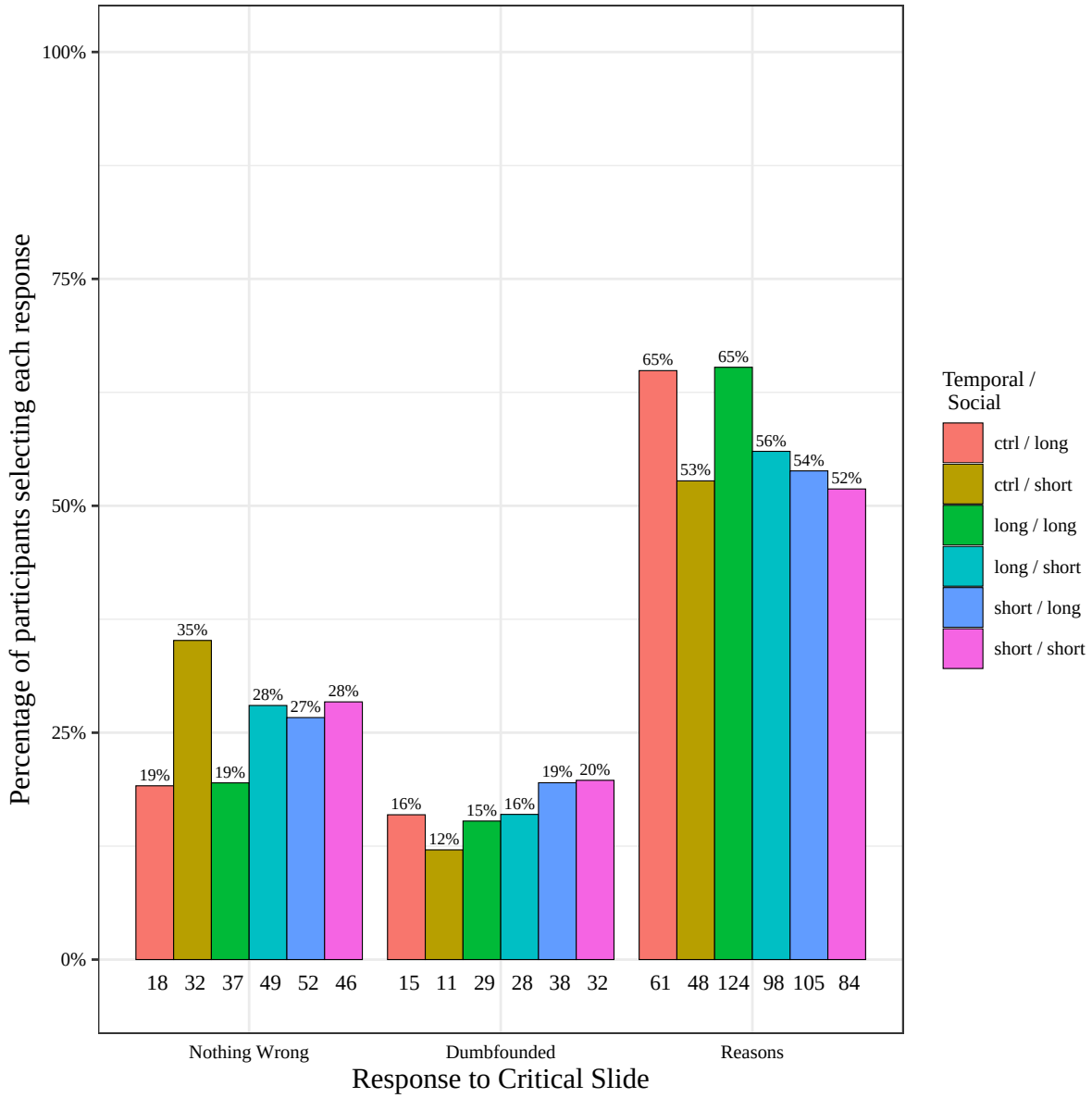
Studies 1, 2, 3, 4, 5, 7, and 8 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on distance manipulation

		ctrl-long	ctrl-short	long-long	long-short	short-long	short-short
Obs	N-W	18.00	32	37	49.00	52.00	46.00
	Dumb	15.00	11	29	28.00	38.00	32.00
	Rsn	61.00	48	124	98.00	105.00	84.00
Exp	N-W	24.25	23.48	49.02	45.15	50.31	41.79
	Dumb	15.86	15.35	32.05	29.52	32.89	27.33
	Rsn	53.89	52.17	108.93	100.33	111.80	92.88
Res	N-W	-1.56	2.15*	-2.24*	0.74	0.31	0.83
	Dumb	-0.25	-1.28	-0.66	-0.34	1.10	1.08
	Rsn	1.57	-0.93	2.49*	-0.40	-1.11	-1.56

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$. Obs = Observed Count; Exp = Expected Count; Res = Standardized Residuals. N-W = Nothing Wrong; Dumb = Dumbfounded; Rsn = Reasons

Figure 21

Response to the critical slide depending in different distancing manipulations across all pilot studies



Combined Results: With Exclusions

Combined Increased Distance vs Reduced Distance (with exclusions)

Our first analysis did not distinguish between temporal and social distance. We created a new variable with three levels, increased distance, decreased distance, and control. A chi-squared test for association revealed a significant a significant association between experimental condition and response to the critical slide: $\chi^2(4,\ N = 626) =$

10.537, $p = .032$, $V = 0.13$, the observed power was 0.84 The responses to the critical slide for the control group ($N = 97$), the increased distance group ($N = 317$) and the decreased distance group ($N = 212$) are displayed in Figure 22. The observed counts, expected counts and standardized residuals are displayed in Table 5.

Figure 22

Studies 1, 2, 3, 4, 5, 7, and 8: Responses to critical slide depending on condition

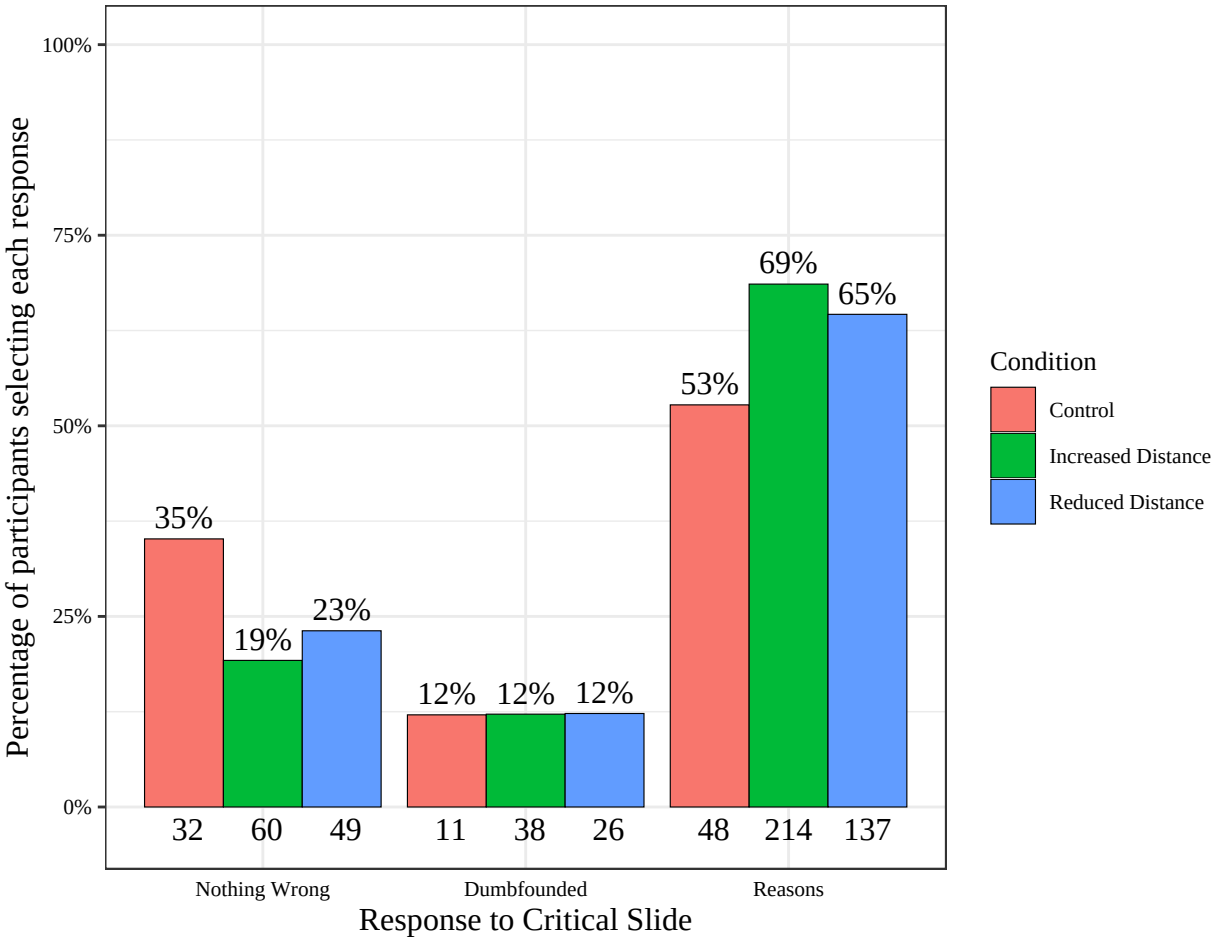


Table 5

Studies 1, 2, 3, 4, 5, 7, and 8 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending distance manipulation

		Control	Long	Short
Observed count	Nothing Wrong	32	60	49.00
	Dumbfounded	11	38	26.00
	Reasons	48	214	137.00
Expected count	Nothing Wrong	20.86	71.53	48.60
	Dumbfounded	11.1	38.05	25.85
	Reasons	59.04	202.42	137.54
Standardised residuals	Nothing Wrong	3.01*	-2.21*	0.08
	Dumbfounded	-0.03	-0.01	0.04
	Reasons	-2.63*	1.96	-0.10

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Different types of manipulations: Temporal / Social (with exclusions)

Overall, there was no significant association between experimental condition and response to the critical slide: $\chi^2(10, N = 615) = 15.753$, $p = .107$, $V = 0.16$, the observed power was 0.95. The responses to the critical slide for each type of manipulation (the total samples are displayed below) are displayed in Figure 23. The observed counts, expected counts and standardized residuals are displayed in Table 6.

Table 6

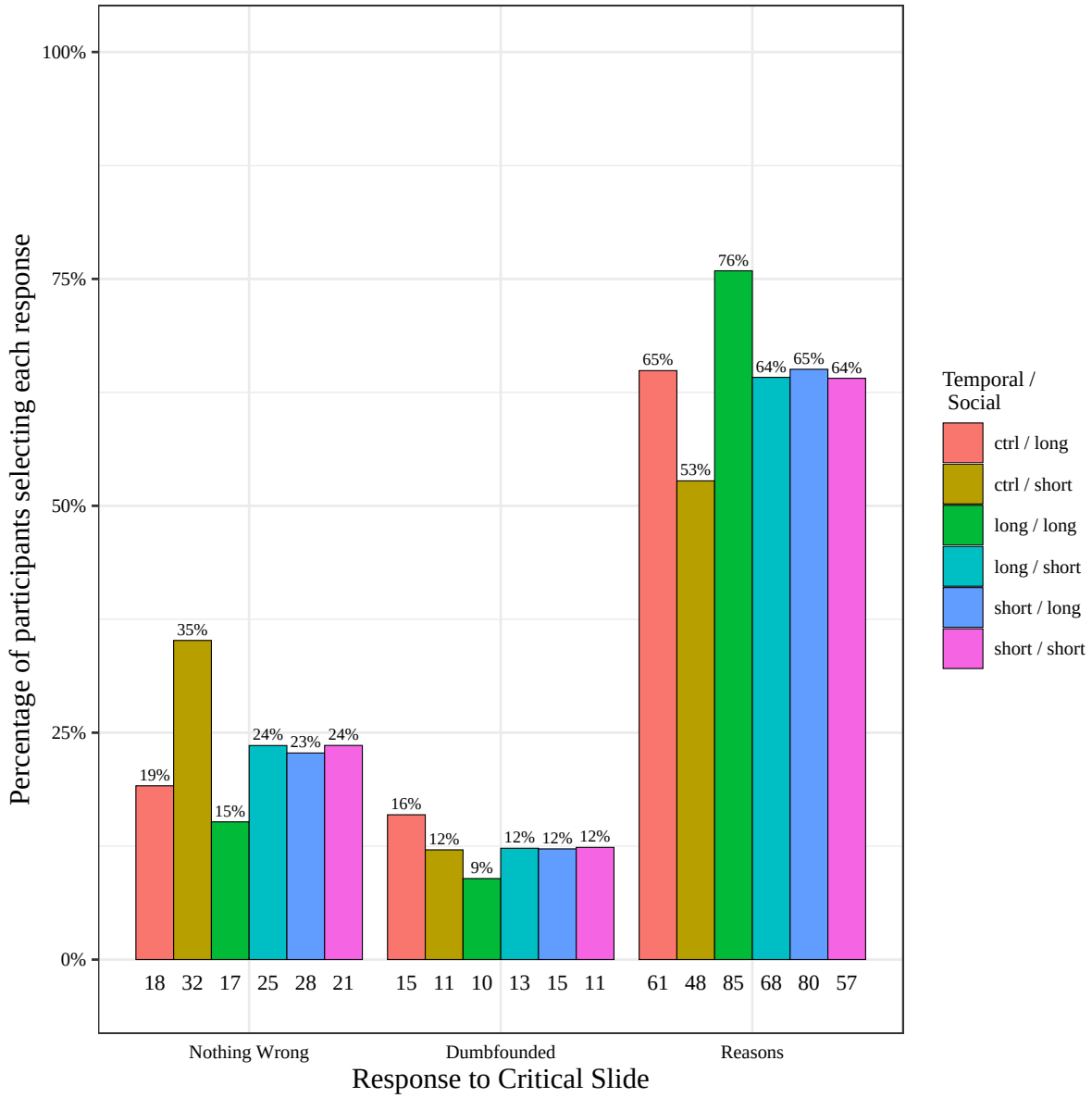
Studies 1, 2, 3, 4, 5, 7, and 8 – Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on distance manipulation

		ctrl-long	ctrl-short	long-long	long-short	short-long	short-short
Obs	N-W	18.00	32	17	25.00	28.00	21.00
	Dumb	15.00	11	10	13.00	15.00	11.00
	Rsn	61.00	48	85	68.00	80.00	57.00
Exp	N-W	21.55	20.86	25.68	24.30	28.20	20.40
	Dumb	11.46	11.1	13.66	12.93	15.00	10.85
	Rsn	60.99	59.04	72.66	68.77	79.80	57.74
Res	N-W	-0.95	3.01*	-2.16*	0.18	-0.05	0.16
	Dumb	1.21	-0.03	-1.17	0.02	0.00	0.05
	Rsn	0.00	-2.63*	2.7*	-0.17	0.04	-0.18

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$. Obs = Observed Count; Exp = Expected Count; Res = Standardized Residuals. N-W = Nothing Wrong; Dumb = Dumbfounded; Rsn = Reasons

Figure 23

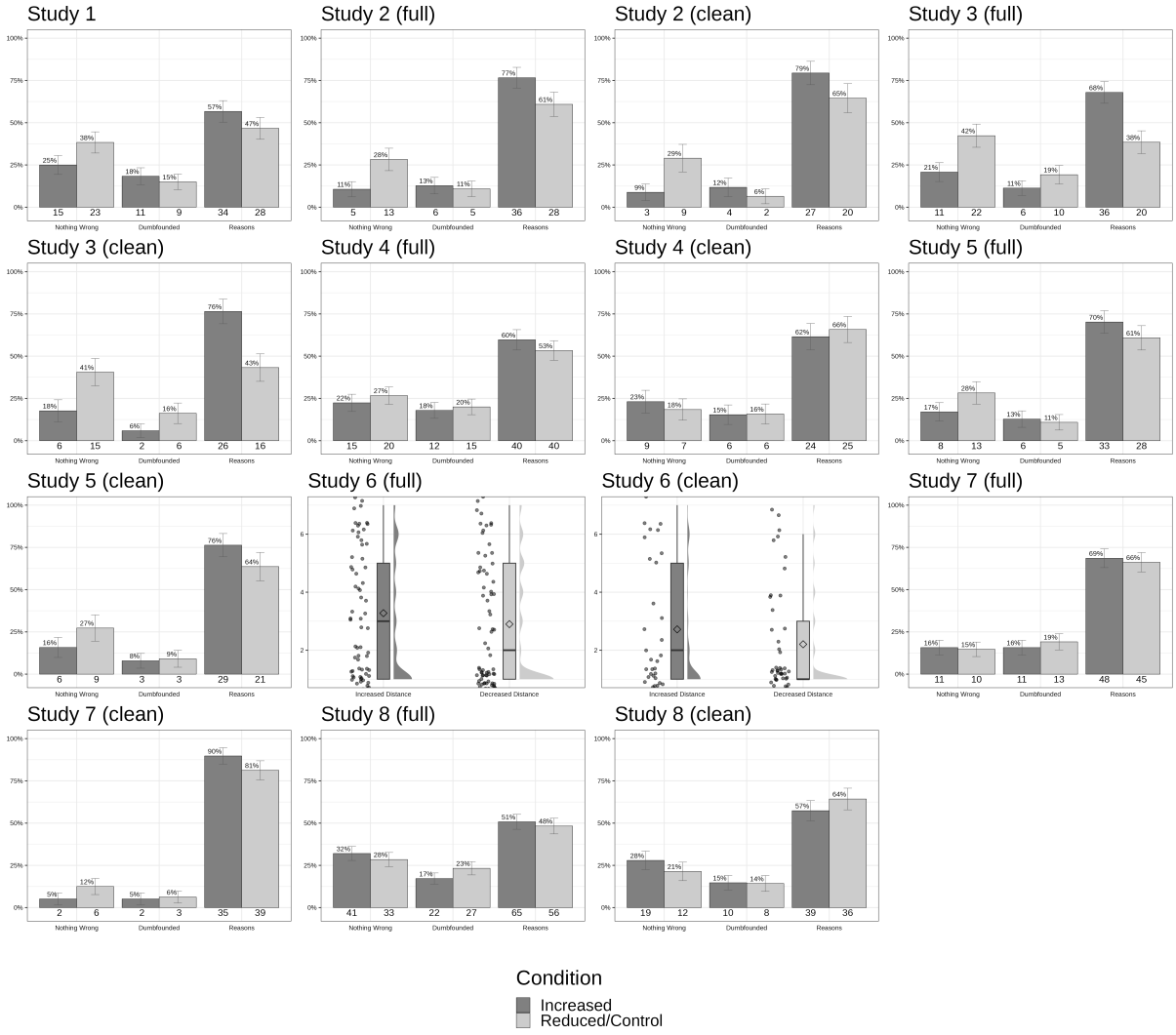
Response to the critical slide depending in different distancing manipulations across all pilot studies



Overview of All Pilot Studies

Figure 24

Results of Studies 1-8.



Analysis of Simulated Data

The following is a sample analysis using a simulated data set. This simulated data set contains $N = 4800$ participants. This sample size was chosen in order for at least 400 participants at each level of the temporal distance IV when analysing the scenarios separately (i.e., aiming for $n = 1200$ per scenario). This sample size will also allow for the detection of a medium effect in the combined analysis on the entire sample.

```
        control future now
other     800    800 800
self      800    800 800
```

Temporal Distancing and Dumbfounding

Overview of Judgments. A total of 3411 participants (71.06%) rated the behavior of Julie and Mark as wrong initially, and 3242 participants (67.54%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings ($M = 3.3$, $SD = 1.3$) and revised ratings ($M = 3.4$, $SD = 1.4$), $t(4799) =$ -3.07, $p = .002$, $d = 0.04$.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $F(2, 4797) = 0.22$, $p = .800$ ${\eta_p}^2 = 0$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.3$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.3$, $M_{\text{control}} = 3.3$, $SD_{\text{control}} = 1.3$). There was no difference in revised judgement depending on distance manipulation: $F(2, 4797) = 0.45$, $p = .640$, ${\eta_p}^2 = 0$, ($M_{\text{increased}} = 3.4$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.4$, $M_{\text{control}} = 3.4$, $SD_{\text{control}} = 1.4$).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(4, N = 4800) = 98.375$, $p < .001$, $V = 0.14$, the observed power was 1. The responses to the critical slide for the increased distance group ($N = 1600$) the decreased distance group ($N = 1600$), and the control group ($N = 1600$) are displayed in Figure 25.

Figure 25

Simulated Data: Responses to critical slide depending on temporal distance for the increased temporal distance group (future, N = 1,600), for the decreased temporal distance group (today, N = 1,600), and for the control group (N = 1,600) (error bars represent standard error of the proportion)

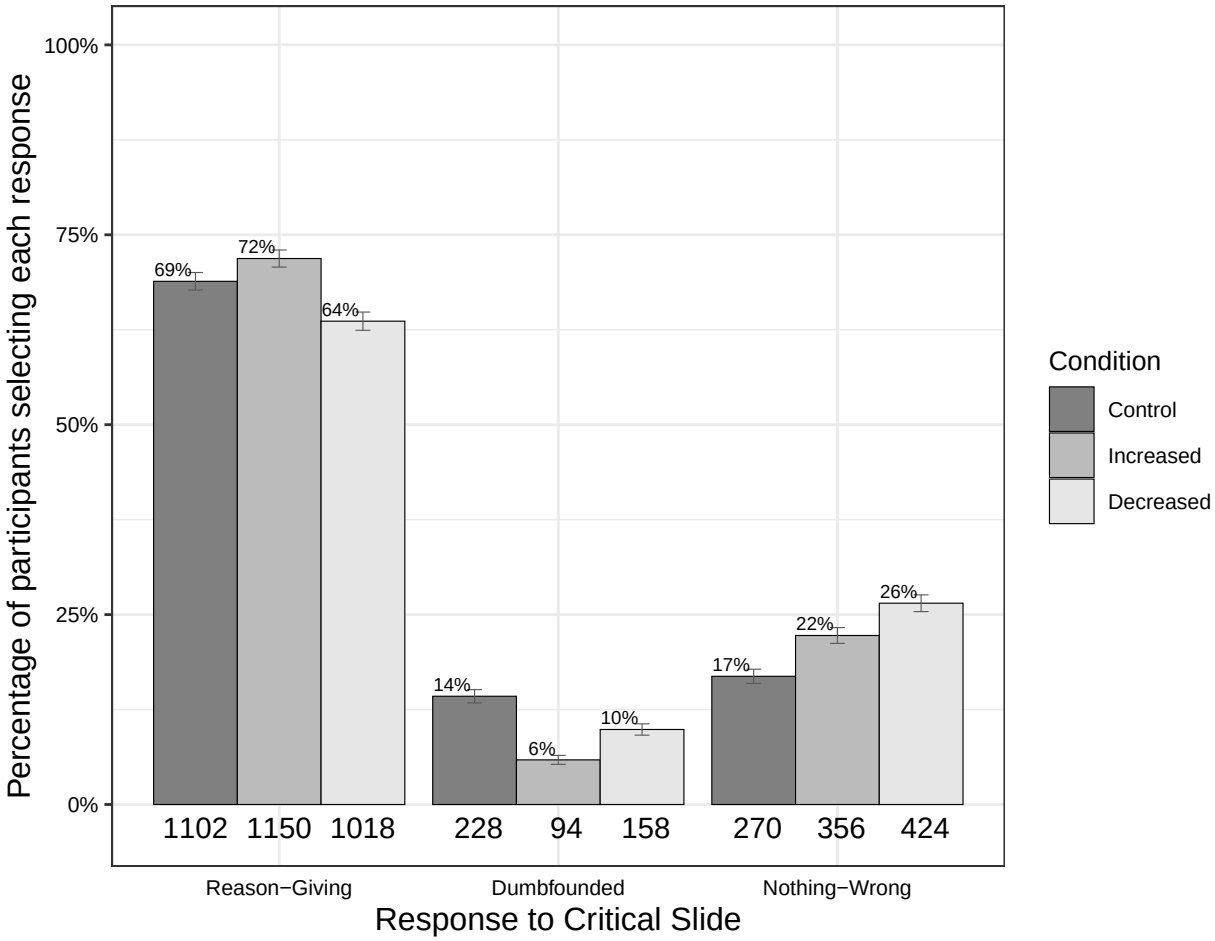


Table 7

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on temporal distancing

		Control	Increased	Decreased
Observed count	Reasons	1102	1150	1018
	Dumbfounded	228	94	158
	Nothing Wrong	270	356	424
Expected count	Reasons	1090	1090	1090
	Dumbfounded	160	160	160
	Nothing Wrong	350	350	350
Standardised residuals	Reasons	0.79	3.94**	-4.73**
	Dumbfounded	6.94**	-6.74**	-0.2
	Nothing Wrong	-5.93**	0.44	5.48**

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Social Distancing and Dumbfounding

Overview of Judgments. A total of 3411 participants (71.06%) rated the behavior of Julie and Mark as wrong initially, and 3242 participants (67.54%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.4, SD = 1.4), $t(4799)$ = -3.07, p = .002, d = 0.04.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $t(4797.07)$ = -1.25, p = .210, d = 0.04, ($M_{\text{increased}}$ = 3.2, $SD_{\text{increased}}$ = 1.3, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.3). There was no difference in revised judgement depending on distance manipulation: $t(4788.04)$ = 1.32, p = .188, d = 0.04, ($M_{\text{increased}}$ = 3.4, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.4).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(2, N = 4800) = 13.911$, $p < .001$, $V = 0.05$, the observed power was 0.89. The responses to the critical slide for the increased distance group (N = 2400) and the decreased distance group (N = 2400) are displayed in Figure 26.

Figure 26

Simulated Data: Responses to critical slide depending on social distance for the increased social distance group (future, N = 2,400), and for the decreased social distance group (today, N = 2,400), (error bars represent standard error of the proportion)

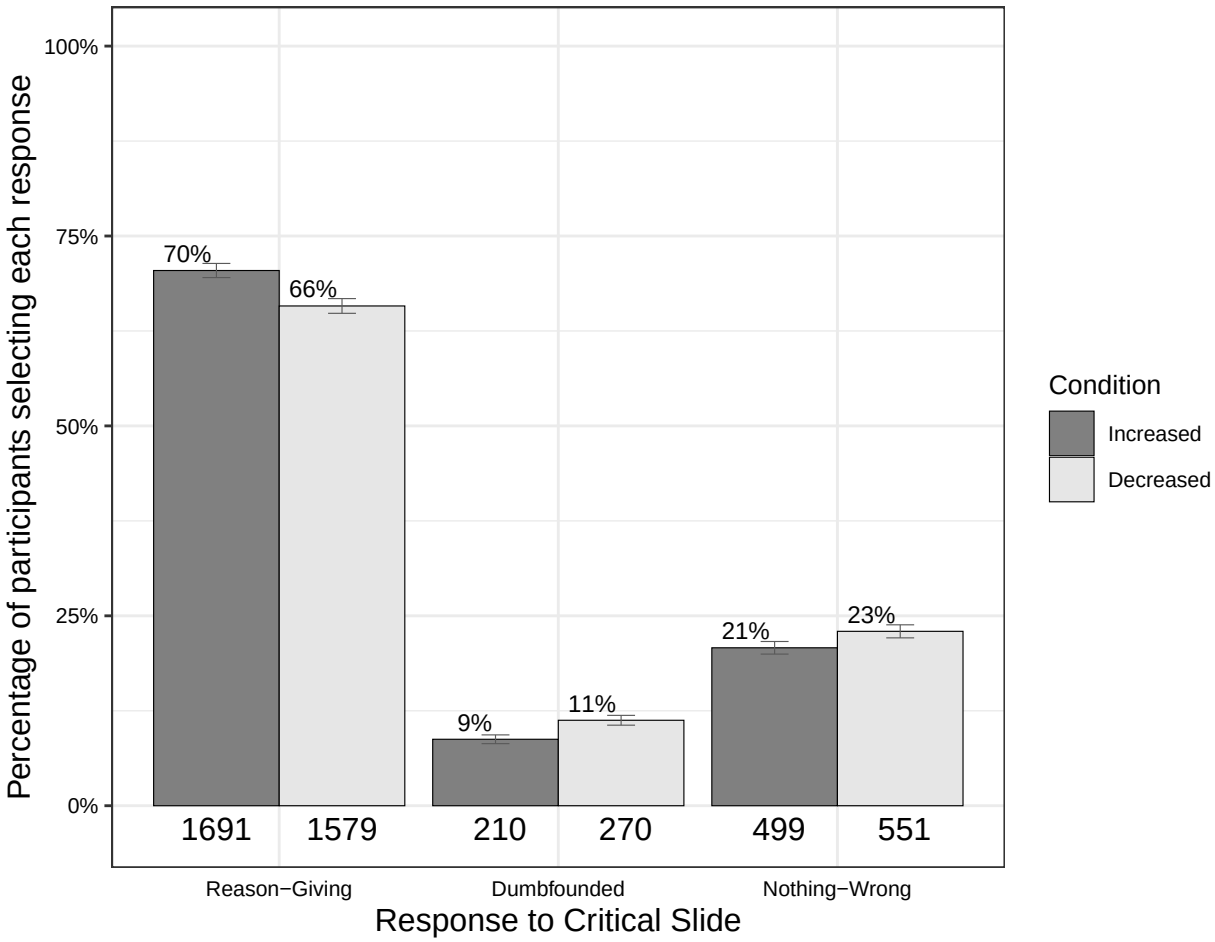


Table 8

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on social distancing

		Increased	Decreased
Observed count	Reasons	1691	1579
	Dumbfounded	210	270
	Nothing Wrong	499	551
Expected count	Reasons	1635	1635
	Dumbfounded	240	240
	Nothing Wrong	525	525
Standardised residuals	Reasons	3.47**	-3.47**
	Dumbfounded	-2.89*	2.89*
	Nothing Wrong	-1.82	1.82

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Distancing and Dumbfounding

Without Scenario

Overall the model significantly predicted responses to the critical slide $\chi^2(10, N = 4800) = 125.47$, $p < .001$, The observed power was 1. The model explained between 0.87% (Cox and Snell R square) and 2.05% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -11.48, $p < .001$, odds ratio = 0.32, 95% CI [0.21, 0.47].

Table 9

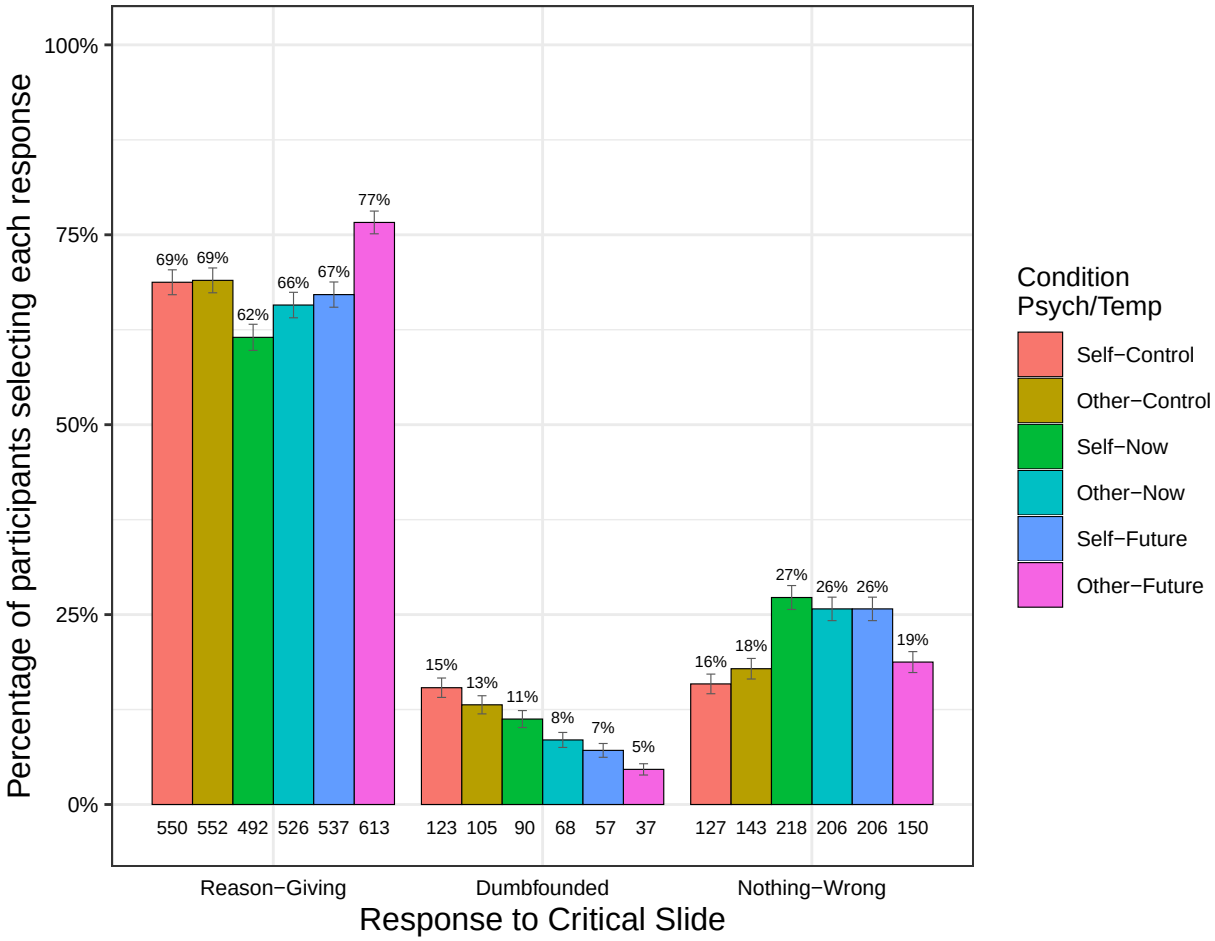
Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.66	0.11	-15.59	-31.18	< .001**	0.19	0.15	0.23
D	Soc-Self	0.16	0.15	1.11	2.22	0.267	1.18	0.88	1.56
D	Temp-Future	-1.15	0.20	-5.74	-11.48	< .001**	0.32	0.21	0.47
D	Temp-Now	-0.39	0.17	-2.31	-4.62	.021*	0.68	0.49	0.94
D	Soc-Self × Temp-Future	0.40	0.26	1.53	3.06	0.126	1.50	0.89	2.51
D	Soc-Self × Temp-Now	0.19	0.23	0.82	1.64	0.412	1.20	0.77	1.87
NW	(Intercept)	-1.35	0.09	-14.39	-28.79	< .001**	0.26	0.22	0.31
NW	Soc-Self	-0.12	0.14	-0.85	-1.69	0.398	0.89	0.68	1.16
NW	Temp-Future	-0.06	0.13	-0.44	-0.87	0.663	0.94	0.73	1.22
NW	Temp-Now	0.41	0.12	3.31	6.63	< .001**	1.51	1.18	1.93
NW	Soc-Self × Temp-Future	0.56	0.18	3.08	6.17	.002*	1.76	1.23	2.52
NW	Soc-Self × Temp-Now	0.24	0.18	1.34	2.67	0.182	1.27	0.89	1.80

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Figure 27

Simulated Data: Responses to critical slide depending on both manipulations. Sample sizes as follows: Self-Control, N = 800, Other-Control, N = 800, Self-Now, N = 800, Other-Now, N = 800, Self-Future, N = 800, Other-Future, N = 800, (error bars represent standard error of the proportion)



Including Scenario

Overall the model significantly predicted responses to the critical slide $\chi^2(16, N = 4800) = 181.64$, $p < .001$, The observed power was 1. The model explained between 1.25% (Cox and Snell R square) and 2.97% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -11.39, $p < .001$, odds ratio = 0.32, 95% CI [0.22, 0.47].

Table 10

Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.69	0.14	-12.15	-24.30	< .001**	0.18	0.14	0.24
D	Soc-Self	0.19	0.15	1.27	2.54	0.204	1.20	0.90	1.61
D	Temp-Future	-1.14	0.20	-5.70	-11.39	< .001**	0.32	0.22	0.47
D	Temp-Now	-0.38	0.17	-2.29	-4.58	.022*	0.68	0.49	0.95
D	Scenario-Jennifer	-0.01	0.14	-0.07	-0.15	0.941	0.99	0.75	1.30
D	Scenario-Julie and Mark	0.36	0.13	2.71	5.42	.007*	1.44	1.11	1.86
D	Scenario-Trolley	-0.41	0.16	-2.62	-5.24	.009*	0.66	0.49	0.90
D	Soc-Self × Temp-Future	0.38	0.26	1.44	2.88	0.15	1.46	0.87	2.45
D	Soc-Self × Temp-Now	0.17	0.23	0.74	1.47	0.462	1.18	0.76	1.84
NW	(Intercept)	-1.32	0.11	-11.73	-23.45	< .001**	0.27	0.21	0.33
NW	Soc-Self	-0.11	0.14	-0.80	-1.60	0.424	0.90	0.69	1.17
NW	Temp-Future	-0.04	0.13	-0.29	-0.59	0.768	0.96	0.74	1.24
NW	Temp-Now	0.44	0.13	3.54	7.08	< .001**	1.56	1.22	1.99
NW	Scenario-Jennifer	-0.36	0.11	-3.40	-6.79	< .001**	0.70	0.57	0.86
NW	Scenario-Julie and Mark	0.06	0.10	0.56	1.12	0.575	1.06	0.87	1.29
NW	Scenario-Trolley	0.11	0.10	1.10	2.20	0.272	1.12	0.92	1.36
NW	Soc-Self × Temp-Future	0.54	0.18	2.96	5.92	.003*	1.72	1.20	2.47
NW	Soc-Self × Temp-Now	0.22	0.18	1.23	2.45	0.22	1.25	0.88	1.77

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Results for Each Scenario

Julie and Mark

Temporal Distancing and Dumbfounding.

Overview of Judgments. A total of 890 participants (70.97%) rated the behavior of Julie and Mark as wrong initially, and 867 participants (69.14%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.4) and revised ratings (M = 3.4, SD = 1.4), $t(1253)$ = -1.36, p = .173, d = 0.04.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $F(2, 1251) = 0.84$, $p = .432$ ${\eta_p}^2$ = 0.001, ($M_{\text{increased}}$ = 3.2, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.4, M_{control} = 3.3, SD_{control} = 1.3). There was no difference in revised judgement depending on distance manipulation: $F(2, 1251) = 0.2$, $p = .818$, ${\eta_p}^2 = 0$, ($M_{\text{increased}}$ = 3.4, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.4, M_{control} = 3.4, SD_{control} = 1.4).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(4, N = 1254) = 30.341$, $p < .001$, $V = 0.16$, the observed power was 1. The responses to the critical slide for the increased distance group (N = 414) the decreased distance group (N = 408), and the control group (N = 432) are displayed in Figure 28.

Table 11

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on temporal distancing

		Control	Increased	Decreased
Observed count	Reasons	293	279	232
	Dumbfounded	70	35	61
	Nothing Wrong	69	100	115
Expected count	Reasons	276.98	265.44	261.59
	Dumbfounded	57.19	54.8	54.01
	Nothing Wrong	97.84	93.76	92.4
Standardised residuals	Reasons	1.99*	1.7	-3.72**
	Dumbfounded	2.25*	-3.51**	1.24
	Nothing Wrong	-4.09**	0.9	3.25*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Social Distancing and Dumbfounding.

Overview of Judgments. A total of 890 participants (70.97%) rated the behavior of Julie and Mark as wrong initially, and 867 participants (69.14%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.4) and revised ratings (M = 3.4, SD = 1.4), $t(1253)$ = -1.36, p = .173, d = 0.04.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $t(1251.88) = 0.59$, $p = .557$, $d = 0.03$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.3$). There was no difference in revised judgement depending on distance manipulation: $t(1250.88) = 0.19$, $p = .850$, $d = 0.01$, ($M_{\text{increased}} = 3.4$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.4$).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(2, N = 1254) = 1.882$, $p = .390$, $V = 0.04$, the observed power was 0.18. The responses to the critical slide for the increased distance group ($N = 633$) and the decreased distance group ($N = 621$) are displayed in Figure 29.

Table 12

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on social distancing

		Increased	Decreased
Observed count	Reasons	399.00	405.00
	Dumbfounded	92.00	74.00
	Nothing Wrong	142.00	142.00
Expected count	Reasons	405.85	398.15
	Dumbfounded	83.79	82.21
	Nothing Wrong	143.36	140.64
Standardised residuals	Reasons	-0.81	0.81
	Dumbfounded	1.37	-1.37
	Nothing Wrong	-0.18	0.18

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Combined Effects of both Distance Manipulations

A multinomial logistic regression was conducted to examine the combined effects of both temporal and social distance on responses to the critical slide. Overall the model significantly predicted responses to the critical slide $\chi^2(10, N = 1254) = 37.31$, $p < .001$, The observed power was 1. The model explained between 0.99% (Cox and Snell R square) and 2.21% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -5.54, $p = .006$, odds ratio = 0.42, 95% CI [0.22, 0.77].

Table 13

Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.30	0.18	-7.36	-14.72	< .001**	0.27	0.19	0.39
D	Soc-Self	-0.30	0.27	-1.11	-2.22	0.268	0.74	0.44	1.26
D	Temp-Future	-0.88	0.32	-2.77	-5.54	.006*	0.42	0.22	0.77
D	Temp-Now	0.17	0.26	0.65	1.30	0.517	1.19	0.71	1.98
D	Soc-Self × Temp-Future	0.49	0.45	1.09	2.19	0.274	1.64	0.68	3.95
D	Soc-Self × Temp-Now	-0.14	0.40	-0.35	-0.69	0.73	0.87	0.40	1.90
NW	(Intercept)	-1.35	0.18	-7.49	-14.99	< .001**	0.26	0.18	0.37
NW	Soc-Self	-0.21	0.27	-0.80	-1.59	0.426	0.81	0.48	1.37
NW	Temp-Future	0.18	0.25	0.73	1.46	0.464	1.20	0.74	1.96
NW	Temp-Now	0.74	0.24	3.07	6.14	.002*	2.10	1.31	3.38
NW	Soc-Self × Temp-Future	0.48	0.36	1.34	2.67	0.181	1.61	0.80	3.25
NW	Soc-Self × Temp-Now	0.02	0.35	0.07	0.14	0.945	1.02	0.51	2.05

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Jennifer

Temporal Distancing and Dumbfounding.

Overview of Judgments. A total of 872 participants (71.24%) rated the behavior of Julie and Mark as wrong initially, and 836 participants (68.3%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.3, SD = 1.4), $t(1223)$ = -0.88, p = .381, d = 0.03.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $F(2, 1221) = 0.11$, $p = .894$ ${\eta_p}^2 = 0$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.3$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.3$, $M_{\text{control}} = 3.3$, $SD_{\text{control}} = 1.3$). There was no difference in revised judgement depending on distance manipulation: $F(2, 1221) = 1.05$, $p = .350$, ${\eta_p}^2 = 0.002$, ($M_{\text{increased}} = 3.4$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.4$, $M_{\text{control}} = 3.3$, $SD_{\text{control}} = 1.4$).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(4, N = 1224) = 42.018$, $p < .001$, $V = 0.19$, the observed power was 1. The responses to the critical slide for the increased distance group (N = 400) the decreased distance group (N = 443), and the control group (N = 381) are displayed in Figure 28.

Table 14

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on temporal distancing

		Control	Increased	Decreased
Observed count	Reasons	284	310	295
	Dumbfounded	58	22	45
	Nothing Wrong	39	68	103
Expected count	Reasons	276.72	290.52	321.75
	Dumbfounded	38.91	40.85	45.24
	Nothing Wrong	65.37	68.63	76
Standardised residuals	Reasons	1.01	2.66*	-3.57**
	Dumbfounded	3.89**	-3.79**	-0.05
	Nothing Wrong	-4.32**	-0.1	4.26**

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Social Distancing and Dumbfounding.

Overview of Judgments. A total of 872 participants (71.24%) rated the behavior of Julie and Mark as wrong initially, and 836 participants (68.3%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.3, SD = 1.4), $t(1223)$ = -0.88, p = .381, d = 0.03.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $t(1215.7)$ = -0.93, p = .354, d = 0.05, ($M_{\text{increased}}$ = 3.2, $SD_{\text{increased}}$ = 1.3, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.3). There was no difference in revised judgement depending on distance manipulation: $t(1221.5)$ = 2.27, p = .023, d = 0.13, ($M_{\text{increased}}$ = 3.4, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.2, $SD_{\text{decreased}}$ = 1.4).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(2, N = 1224) = 20.849$, $p < .001$, $V = 0.13$, the observed power was 0.98. The responses to the critical slide for the increased distance group (N = 632) and the decreased distance group (N = 592) are displayed in Figure 29.

Table 15

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on social distancing

		Increased	Decreased
Observed count	Reasons	488	401
	Dumbfounded	42	83
	Nothing Wrong	102	108
Expected count	Reasons	459.03	429.97
	Dumbfounded	64.54	60.46
	Nothing Wrong	108.43	101.57
Standardised residuals	Reasons	3.72**	-3.72**
	Dumbfounded	-4.26**	4.26**
	Nothing Wrong	-0.98	0.98

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Combined Effects of both Distance Manipulations

A multinomial logistic regression was conducted to examine the combined effects of both temporal and social distance on responses to the critical slide. Overall the model significantly predicted responses to the critical slide $\chi^2(10, N = 1224) = 65.99$, $p < .001$, The observed power was 1. The model explained between 1.78% (Cox and Snell R square) and 4.45% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -4.89, $p = .015$, odds ratio = 0.36, 95% CI [0.16, 0.82].

Table 16

Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.96	0.24	-8.21	-16.41	< .001**	0.14	0.09	0.22
D	Soc-Self	0.64	0.30	2.13	4.27	.033*	1.90	1.05	3.42
D	Temp-Future	-1.02	0.42	-2.44	-4.89	.015*	0.36	0.16	0.82
D	Temp-Now	-0.60	0.37	-1.62	-3.23	0.106	0.55	0.26	1.14
D	Soc-Self × Temp-Future	0.01	0.54	0.02	0.04	0.983	1.01	0.35	2.91
D	Soc-Self × Temp-Now	0.55	0.46	1.20	2.40	0.231	1.74	0.70	4.29
NW	(Intercept)	-2.01	0.24	-8.23	-16.47	< .001**	0.13	0.08	0.22
NW	Soc-Self	0.05	0.34	0.15	0.30	0.88	1.05	0.54	2.06
NW	Temp-Future	0.27	0.31	0.86	1.72	0.389	1.31	0.71	2.41
NW	Temp-Now	0.83	0.29	2.86	5.72	.004*	2.30	1.30	4.07
NW	Soc-Self × Temp-Future	0.41	0.43	0.95	1.89	0.344	1.51	0.64	3.54
NW	Soc-Self × Temp-Now	0.22	0.41	0.54	1.08	0.588	1.25	0.56	2.80

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Trolley

Temporal Distancing and Dumbfounding.

Overview of Judgments. A total of 826 participants (70.96%) rated the behavior of Julie and Mark as wrong initially, and 773 participants (66.41%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.4, SD = 1.4), $t(1163)$ = -2.15, p = .032, d = 0.06.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $F(2, 1161) = 1.51$, $p = .220$ ${\eta_p}^2 = 0.003$, ($M_{\text{increased}} = 3.4$, $SD_{\text{increased}} = 1.3$, $M_{\text{decreased}} = 3.2$, $SD_{\text{decreased}} = 1.3$, $M_{\text{control}} = 3.2$, $SD_{\text{control}} = 1.3$). There was no difference in revised judgement depending on distance manipulation: $F(2, 1161) = 0.6$, $p = .546$, ${\eta_p}^2 = 0.001$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.3$, $M_{\text{decreased}} = 3.4$, $SD_{\text{decreased}} = 1.4$, $M_{\text{control}} = 3.4$, $SD_{\text{control}} = 1.4$).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(4, N = 1164) = 33.49$, $p < .001$, $V = 0.17$, the observed power was 1. The responses to the critical slide for the increased distance group (N = 393) the decreased distance group (N = 360), and the control group (N = 411) are displayed in Figure 28.

Table 17

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on temporal distancing

		Control	Increased	Decreased
Observed count	Reasons	271	285	238
	Dumbfounded	49	13	15
	Nothing Wrong	91	95	107
Expected count	Reasons	280.36	268.08	245.57
	Dumbfounded	27.19	26	23.81
	Nothing Wrong	103.46	98.93	90.62
Standardised residuals	Reasons	-1.23	2.25*	-1.03
	Dumbfounded	5.38**	-3.24*	-2.25*
	Nothing Wrong	-1.76	-0.56	2.39*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Social Distancing and Dumbfounding.

Overview of Judgments. A total of 826 participants (70.96%) rated the behavior of Julie and Mark as wrong initially, and 773 participants (66.41%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.4, SD = 1.4), $t(1163)$ = -2.15, p = .032, d = 0.06.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $t(1155.77)$ = -0.26, p = .794, d = 0.02, ($M_{\text{increased}}$ = 3.3, $SD_{\text{increased}}$ = 1.3, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.3). There was no difference in revised judgement depending on distance manipulation: $t(1139.87)$ = -0.25, p = .806, d = 0.01, ($M_{\text{increased}}$ = 3.4, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.4, $SD_{\text{decreased}}$ = 1.3).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(2, N = 1164) = 8.948$, p = .011, V = 0.09, the observed power was 0.71. The responses to the critical slide for the increased distance group (N = 556) and the decreased distance group (N = 608) are displayed in Figure 29.

Table 18

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on social distancing

		Increased	Decreased
Observed count	Reasons	488	401
	Dumbfounded	42	83
	Nothing Wrong	102	108
Expected count	Reasons	459.03	429.97
	Dumbfounded	64.54	60.46
	Nothing Wrong	108.43	101.57
Standardised residuals	Reasons	3.72**	-3.72**
	Dumbfounded	-4.26**	4.26**
	Nothing Wrong	-0.98	0.98

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Combined Effects of both Distance Manipulations

A multinomial logistic regression was conducted to examine the combined effects of both temporal and social distance on responses to the critical slide. Overall the model significantly predicted responses to the critical slide $\chi^2(10, N = 1164) = 47.97$, $p < .001$, The observed power was 1. The model explained between 1.36% (Cox and Snell R square) and 3.34% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -7.15, $p < .001$, odds ratio = 0.11, 95% CI [0.03, 0.37].

Table 19

Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.69	0.22	-7.60	-15.21	< .001**	0.18	0.12	0.29
D	Soc-Self	-0.04	0.31	-0.13	-0.26	0.896	0.96	0.52	1.77
D	Temp-Future	-2.23	0.62	-3.57	-7.15	< .001**	0.11	0.03	0.37
D	Temp-Now	-1.51	0.51	-2.97	-5.93	.003*	0.22	0.08	0.60
D	Soc-Self × Temp-Future	1.36	0.74	1.85	3.70	0.064	3.91	0.92	16.60
D	Soc-Self × Temp-Now	0.78	0.64	1.22	2.44	0.223	2.19	0.62	7.72
NW	(Intercept)	-1.13	0.18	-6.37	-12.73	< .001**	0.32	0.23	0.46
NW	Soc-Self	0.07	0.24	0.30	0.60	0.764	1.08	0.67	1.73
NW	Temp-Future	-0.36	0.26	-1.39	-2.78	0.165	0.70	0.42	1.16
NW	Temp-Now	0.13	0.25	0.53	1.06	0.594	1.14	0.70	1.86
NW	Soc-Self × Temp-Future	0.63	0.34	1.83	3.66	0.067	1.88	0.96	3.69
NW	Soc-Self × Temp-Now	0.30	0.34	0.88	1.76	0.378	1.35	0.69	2.61

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Heinz

Temporal Distancing and Dumbfounding.

Overview of Judgments. A total of 823 participants (71.07%) rated the behavior of Julie and Mark as wrong initially, and 766 participants (66.15%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.4, SD = 1.4), $t(1157)$ = -1.81, p = .071, d = 0.05.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $F(2, 1155) = 0.2$, $p = .817$ ${\eta_p}^2 = 0$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.2$, $SD_{\text{decreased}} = 1.3$, $M_{\text{control}} = 3.3$, $SD_{\text{control}} = 1.3$). There was no difference in revised judgement depending on distance manipulation: $F(2, 1155) = 1.33$, $p = .264$, ${\eta_p}^2 = 0.002$, ($M_{\text{increased}} = 3.3$, $SD_{\text{increased}} = 1.4$, $M_{\text{decreased}} = 3.3$, $SD_{\text{decreased}} = 1.4$, $M_{\text{control}} = 3.5$, $SD_{\text{control}} = 1.4$).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(4, N = 1158) = 15.71$, $p = .003$, $V = 0.12$, the observed power was 0.93. The responses to the critical slide for the increased distance group (N = 393) the decreased distance group (N = 389), and the control group (N = 376) are displayed in Figure 28.

Table 20

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on temporal distancing

		Control	Increased	Decreased
Observed count	Reasons	271	285	238
	Dumbfounded	49	13	15
	Nothing Wrong	91	95	107
Expected count	Reasons	280.36	268.08	245.57
	Dumbfounded	27.19	26	23.81
	Nothing Wrong	103.46	98.93	90.62
Standardised residuals	Reasons	-1.23	2.25*	-1.03
	Dumbfounded	5.38**	-3.24*	-2.25*
	Nothing Wrong	-1.76	-0.56	2.39*

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Social Distancing and Dumbfounding.

Overview of Judgments. A total of 823 participants (71.07%) rated the behavior of Julie and Mark as wrong initially, and 766 participants (66.15%) rated the behavior as wrong at the end of the task. There was a significant difference between initial ratings (M = 3.3, SD = 1.3) and revised ratings (M = 3.4, SD = 1.4), $t(1157)$ = -1.81, p = .071, d = 0.05.

Distancing and Judgments Made. There was no difference in initial judgement depending on distance manipulation: $t(1153.29)$ = -1.98, p = .048, d = 0.12, ($M_{\text{increased}}$ = 3.2, $SD_{\text{increased}}$ = 1.3, $M_{\text{decreased}}$ = 3.3, $SD_{\text{decreased}}$ = 1.4). There was no difference in revised judgement depending on distance manipulation: $t(1143.18)$ = 0.44, p = .661, d = 0.03, ($M_{\text{increased}}$ = 3.4, $SD_{\text{increased}}$ = 1.4, $M_{\text{decreased}}$ = 3.4, $SD_{\text{decreased}}$ = 1.3).

Distancing and Reason-Giving/Dumbfounding. There was a significant association between temporal distance condition and response to the critical slide, $\chi^2(2, N = 1158) = 5.699$, p = .058, V = 0.07, the observed power was 0.5. The responses to the critical slide for the increased distance group (N = 579) and the decreased distance group (N = 579) are displayed in Figure 29.

Table 21

Observed counts, expected counts, and standardised residuals for each response to the critical slide depending on social distancing

		Increased	Decreased
Observed count	Reasons	488	401
	Dumbfounded	42	83
	Nothing Wrong	102	108
Expected count	Reasons	459.03	429.97
	Dumbfounded	64.54	60.46
	Nothing Wrong	108.43	101.57
Standardised residuals	Reasons	3.72**	-3.72**
	Dumbfounded	-4.26**	4.26**
	Nothing Wrong	-0.98	0.98

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$

Combined Effects of both Distance Manipulations

A multinomial logistic regression was conducted to examine the combined effects of both temporal and social distance on responses to the critical slide. Overall the model significantly predicted responses to the critical slide $\chi^2(10, N = 1158) = 25.95$, $p = .004$, The observed power was 0.96. The model explained between 0.74% (Cox and Snell R square) and 1.76% (Nadelkerke R squared) of the variance in responses to the critical slide. For scenarios in the future, participants were more likely to provide reasons than to present as dumbfounded Wald = -4.29, $p = .032$, odds ratio = 0.41, 95% CI [0.18, 0.93].

Table 22

Predictors of each response with reason-giving as the reference response

Response	Term	*b*	*S.E.*	*z*	Wald	*p*	*O.R.*	Lower	Upper
D	(Intercept)	-1.87	0.24	-7.79	-15.59	< .001**	0.15	0.10	0.25
D	Soc-Self	0.49	0.31	1.55	3.10	0.121	1.63	0.88	3.00
D	Temp-Future	-0.90	0.42	-2.15	-4.29	.032*	0.41	0.18	0.93
D	Temp-Now	-0.26	0.36	-0.73	-1.45	0.468	0.77	0.38	1.57
D	Soc-Self × Temp-Future	0.11	0.54	0.21	0.41	0.837	1.12	0.39	3.21
D	Soc-Self × Temp-Now	-0.10	0.48	-0.20	-0.40	0.841	0.91	0.36	2.31
NW	(Intercept)	-1.11	0.18	-6.29	-12.58	< .001**	0.33	0.23	0.47
NW	Soc-Self	-0.38	0.27	-1.39	-2.78	0.164	0.68	0.40	1.17
NW	Temp-Future	-0.15	0.25	-0.60	-1.19	0.551	0.86	0.53	1.41
NW	Temp-Now	0.17	0.24	0.72	1.44	0.472	1.19	0.74	1.92
NW	Soc-Self × Temp-Future	0.70	0.36	1.93	3.87	0.053	2.02	0.99	4.13
NW	Soc-Self × Temp-Now	0.37	0.36	1.02	2.04	0.309	1.45	0.71	2.94

Note. * = sig. at $p < .05$; ** = sig. at $p < .001$; D = dumbfounded, NW = nothing-wrong

Plots: All Scenarios

Figure 28

Temporal distance and reason-giving for each Scenario

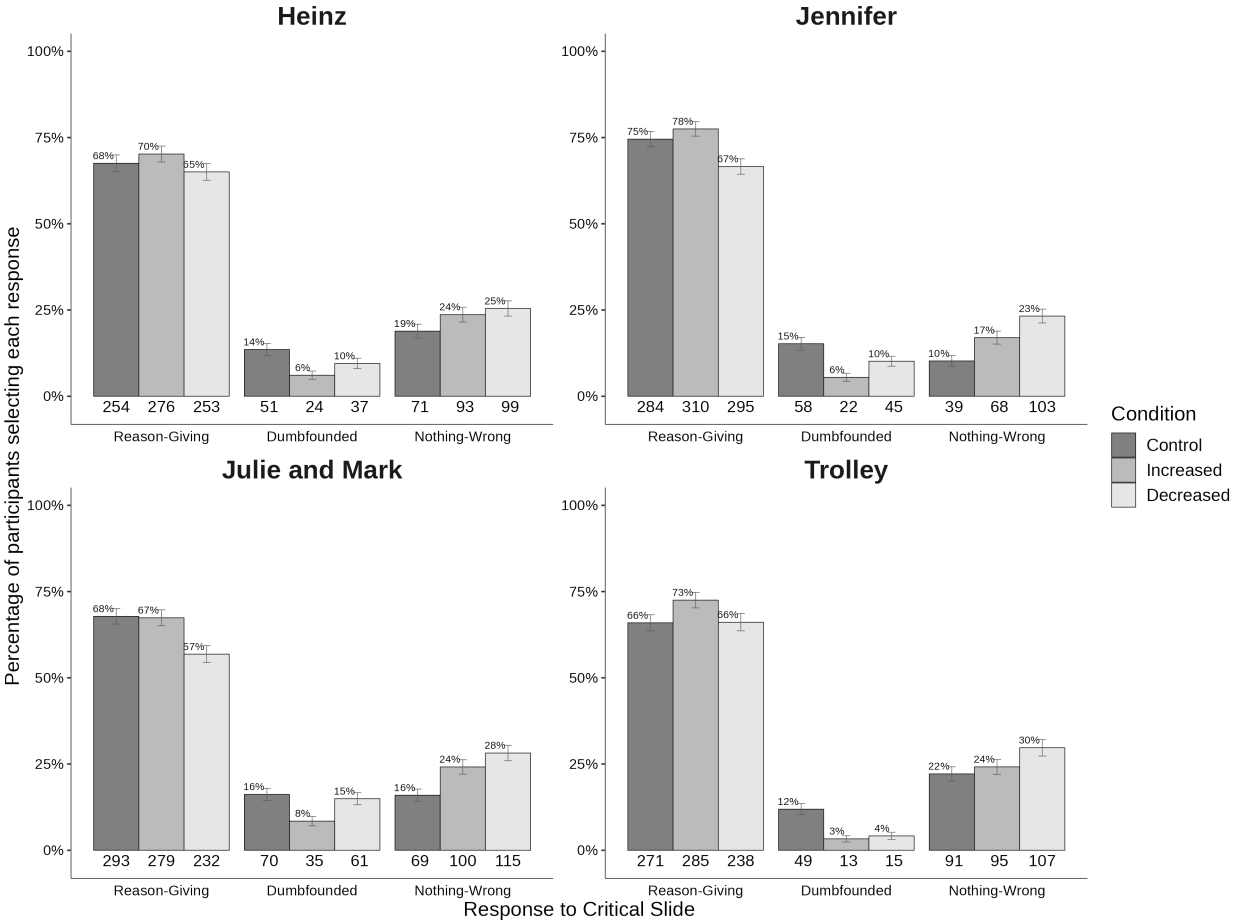


Figure 29

Social distance and reason-giving for each Scenario

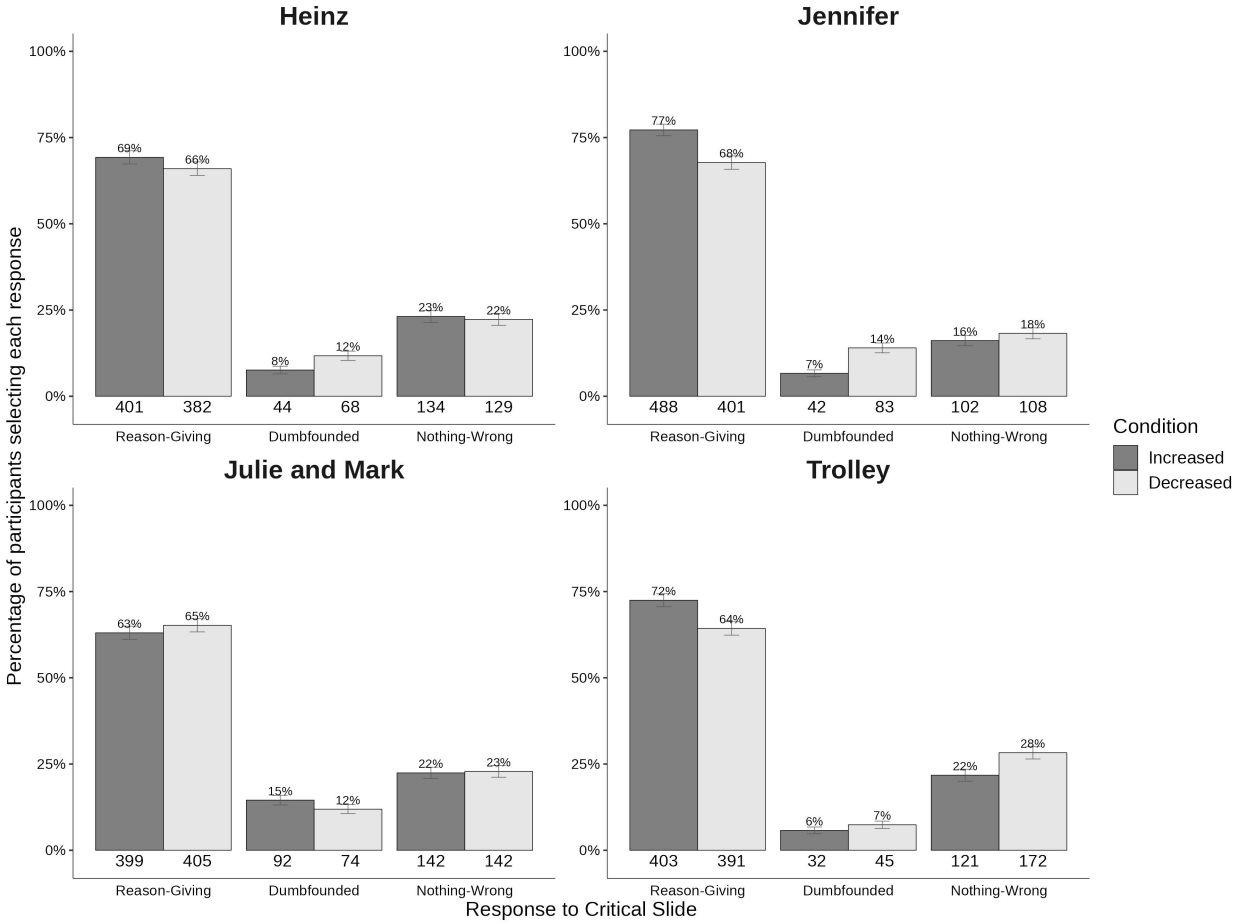
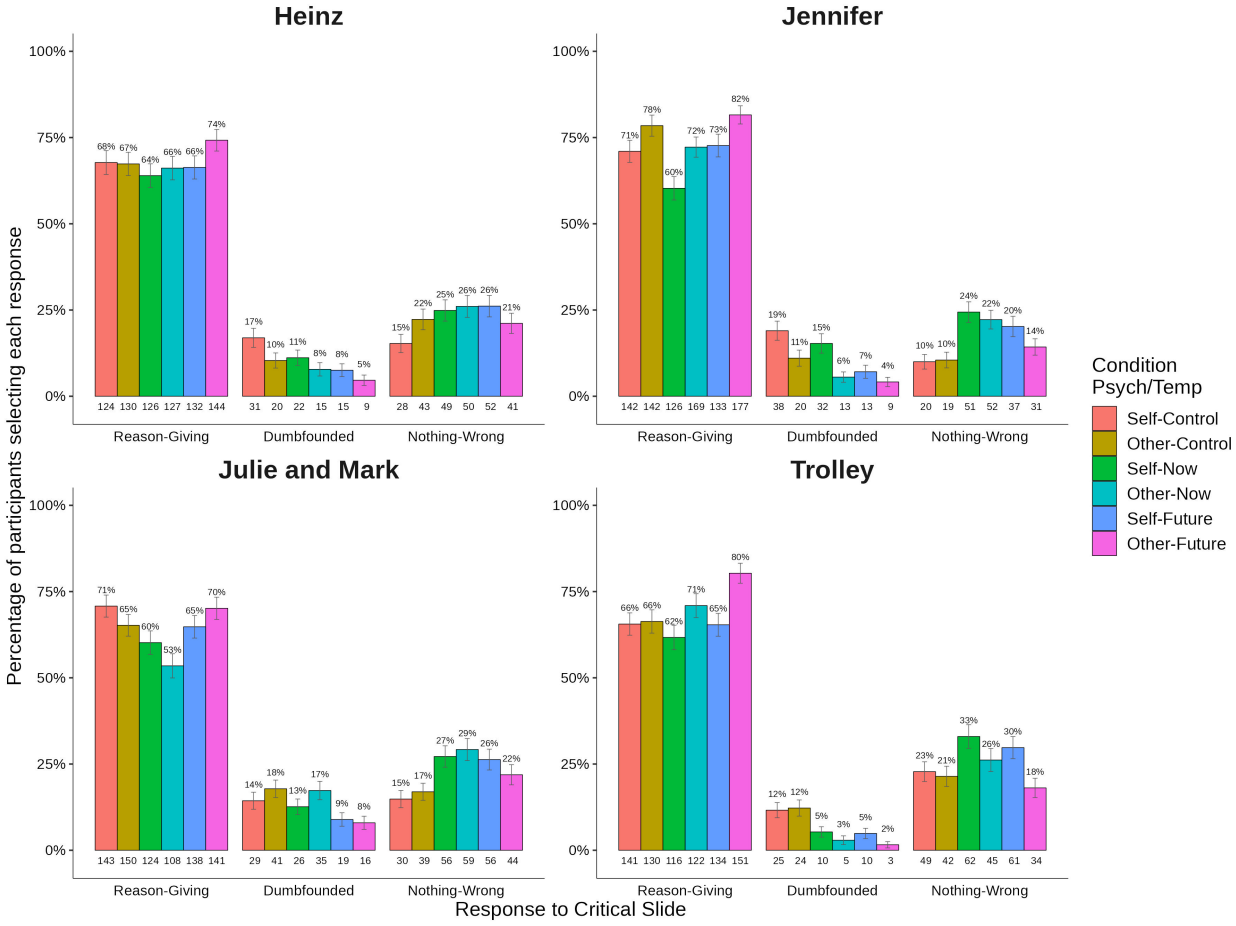


Figure 30

Both distance manipulations and reason-giving for each Scenario



References

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801. https://doi.org/10.1037/xge0000533

Ballard, R. (1992). Short Forms of the Marlowe-Crowne Social Desirability Scale. *Psychological Reports*, *71*(3_suppl), 1155–1160. https://doi.org/10.2466/pr0.1992.71.3f.1155

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition*, *38*(2), 186–196. https://doi.org/10.3758/MC.38.2.186

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. https://doi.org/10.1037/h0047358

Cushman, F. A. (2013). Action, Outcome, and Value A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: Consequences for insight and creative cognition. *Journal of Personality and Social Psychology*, 177–189.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Greene, J. D. (2008). The Secret Joke of Kant's Soul. In *Moral Psychology Volume 3: The neurosciences of morality: Emotion, brain disorders, and development* (pp. 35–79). the MIT press.

Haidt, J., Björklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished Manuscript, University of Virginia.*

Kruglanski, A. W. (1996). Motivated social cognition: Principles of the interface. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 493–520). Guilford Press.

Kruglanski, A. W. (2013). *The Psychology of Closed Mindedness*. Psychology Press.

Kruglanski, A. W., Atash, M. N., De Grada, E., Mannetti, L., & Pierro, A. (2013). *Need for Closure Scale (NFC). Measurement Instrument Database for the Social Science*.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0020762

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, *38*(6), 523–534. https://doi.org/10.1016/S0022-1031(02)00535-8

McHugh, C., Griffin, S. M., McGrath, M. J., Rhee, J. J., Maher, P. J., McCashin, D., & Roth, J. (2022). Moral Identity Predicts Adherence to COVID-19 Mitigation Procedures Depending on Political Ideology: A Comparison Between the USA and New Zealand. *Political Psychology*, *n/a*(n/a). https://doi.org/10.1111/pops.12838

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2017). Searching for Dumbfounding. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/WM6VC

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2020). Reasons or rationalizations: The role of principles in the moral dumbfounding paradigm. *Journal of Behavioral Decision Making*, *33*(3), 376–392. https://doi.org/10.1002/bdm.2167

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2023). Cognitive Load Can Reduce Reason-Giving in a Moral Dumbfounding Task. *Collabra: Psychology*, *9*(1), 73818. https://doi.org/10.1525/collabra.73818

Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4), 296–313.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *28*(2),

191–193. https://doi.org/10.1002/1097-4679(197204)28:2%3C191::AID-JCLP2270280220%3E3.0.CO;2-G

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275. https://doi.org/10.3758/s13421-011-0104-1

Unipark, Q. (2013). *QuestBack Unipark.(2013)*.