The Moral Dilution Effect: Irrelevant Information Influences Judgments of Moral Character

Cillian McHugh[1] & Eric R. Igou[1]

[1] University of Limerick

Author Note

Correspondence concerning this article should be addressed to Cillian McHugh, University of Limerick, Limerick, Ireland, V94 T9PX. E-mail: cillian.mchugh@ul.ie

Abstract

Across five studies we investigated the moral dilution effect

Keywords: keywords

Word count: TBC

The Moral Dilution Effect: Irrelevant Information Influences Judgments of Moral Character

Recent developments in the psychology of moral judgments propose that making a moral judgment is an act of categorization (McHugh, McGann, Igou, & Kinsella, 2022). A corollary of this view is that moral judgments vary depending on how well a target maps onto a template of a prototypical morally relevant act/actor (Schein & Gray, 2018)

Pilot Study 1

The aim of this pilot study was to develop and test materials that could be used to study the dilution effect for moral characters. We developed diagnostic and non-diagnostic character descriptions. We hypothesized that moral evaluations of the diagnostic descriptions would be more severe (more immoral) than for the non-diagnostic descriptions.

Pilot Study 1: Method

Pilot 1: Participants and design

The pilot study was a within-subjects design. The independent variable was description type with two levels, *diagnostic* and *non-diagnostic*. We used two dependent variables. The first dependent variable was the four item moral perception scale (MPS-4), participants rated the characters on four dimensions using 7-point bipolar scales. The dimensions and scale endpoints were: Bad-Good, Immoral-Moral, Violent-Peaceful, Merciless-Empathetic, this showed excellent reliability, $\alpha = 0.93$. The second dependent variable was a single item moral perception measure (MM-1) which consisted of a 100-point slider ranging from $0 = $ *Very Immoral* to $100 = $ *Very Moral*. Both dependent variables were taken from Walker, Turpin, Fugelsang, and Białek (2021).

A total sample of 235 (89 female, 142 male, 1 non-binary, 1 prefer not to say; $M_\text{age} = 36.45$, min $= 20$, max $= 72$, $SD = 10.23$) started the survey. Participants were recruited from MTurk.

We removed participants who failed both manipulation checks ($n = 23$), leaving a total sample of 0 participants (80 female, 128 male, 1 non-binary, 1 prefer not to say; $M_{\text{age}} = 36.63$, min = 20, max = 72, $SD = 10.34$).

Pilot 1: Procedure and materials

Data were collected using an online questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented with descriptions of six characters.

Moral character descriptions were developed by combining descriptions relating to three different moral foundations. These descriptions were adapted from the items of the extended character morality questionnaire (Grizzard et al., 2020). A sample description reads: *Imagine a person named Sam. Throughout their life they have been known to be cruel, act unfairly, and to betray their own group.* Full text of these descriptions can be found in the supplementary materials.

We developed neutral descriptions that included information relating to physical appearance/attributes, hobbies/activities, and family information, e.g., *Imagine a person named Jackie. They have red hair, play tennis four times a month, and have one older sibling and one younger sibling.*

Character descriptions did not specify the gender of the charcters, and all characters had names that could be either male or female (Sam, Robin, Francis, Alex, Jackie, Charlie). All participants read six descriptions, four moral descriptions and two neutral. Pilot Study 1 was pre-registered at https://aspredicted.org/3VK_8FD.

Pilot 1: Results

The means and standard deviations for MPS-4 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MPS-4}} = 4.35$, $SD_{\text{MPS-4}} = 1.90$, *Francis* (diagnostic/moral), $M_{\text{MPS-4}} = 4.46$, $SD_{\text{MPS-4}} = 1.73$, *Alex* (diagnostic/moral), $M_{\text{MPS-4}} = 4.44$, $SD_{\text{MPS-4}} = 1.79$, *Robin* (diagnostic/moral), $M_{\text{MPS-4}} = 4.35$, $SD_{\text{MPS-4}} = 1.96$, *Jackie* (non-diagnostic/neutral),

$M_{\text{MPS-4}} = 5.40$, $SD_{\text{MPS-4}} = 1.01$, *Charlie* (non-diagnostic/neutral), $M_{\text{MPS-4}} = 5.38$, $SD_{\text{MPS-4}} = 1.01$. For the diagnostic descriptions, there was no significant variation depending on the description, $F(599.98, 2.84) = 1.58$, $p = .194$, partial $\eta^2 = 0.00$. For the non-diagnostic descriptions there was no significant difference in ratings depending on description, $t(211) = -0.67$, $p = .506$, $d = 0.05$.

The means and standard deviations for MM-1 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MM-1}} = 55.67$, $SD_{\text{MM-1}} = 30.47$; *Francis* (diagnostic/moral), $M_{\text{MM-1}} = 58.22$, $SD_{\text{MM-1}} = 28.61$; *Alex* (diagnostic/moral), $M_{\text{MM-1}} = 56.80$, $SD_{\text{MM-1}} = 29.45$; *Robin* (diagnostic/moral), $M_{\text{MM-1}} = 55.49$, $SD_{\text{MM-1}} = 31.38$; *Jackie* (non-diagnostic/neutral), $M_{\text{MM-1}} = 73.00$, $SD_{\text{MM-1}} = 14.72$; *Charlie* (non-diagnostic/neutral), $M_{\text{MM-1}} = 72.94$, $SD_{\text{MM-1}} = 14.79$. For the diagnostic descriptions, we observed significant variation depending on the description, $F(608.49, 2.88) = 3.01$, $p = .032$, partial $\eta^2 = 0.001$. When correcting for multiple comparisons, pairwise comparisons did not reveal significant differences between descriptions. We note that without correction, *Francis* appeared to be rated as more moral than both *Robin* ($p = .012$), and *Sam* ($p = .009$). For the non-diagnostic descriptions there was no significant difference in ratings depending on description, $t(211) = -0.09$, $p = .929$, $d = 0.01$.

We conducted a linear-mixed-effects model to test if condition influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(2) = 860.16$, $p < .001$. Condition was a significant predictor in the model $b = -0.49$, $t(211.05) = -8.54$, $p < .001$, with the non-diagnostic (neutral) descriptions being rated as more moral than the diagnostic (morally relevant) descriptions of immoral characters Figure 1.

We conducted a linear-mixed-effects model to test if condition influenced MM-1
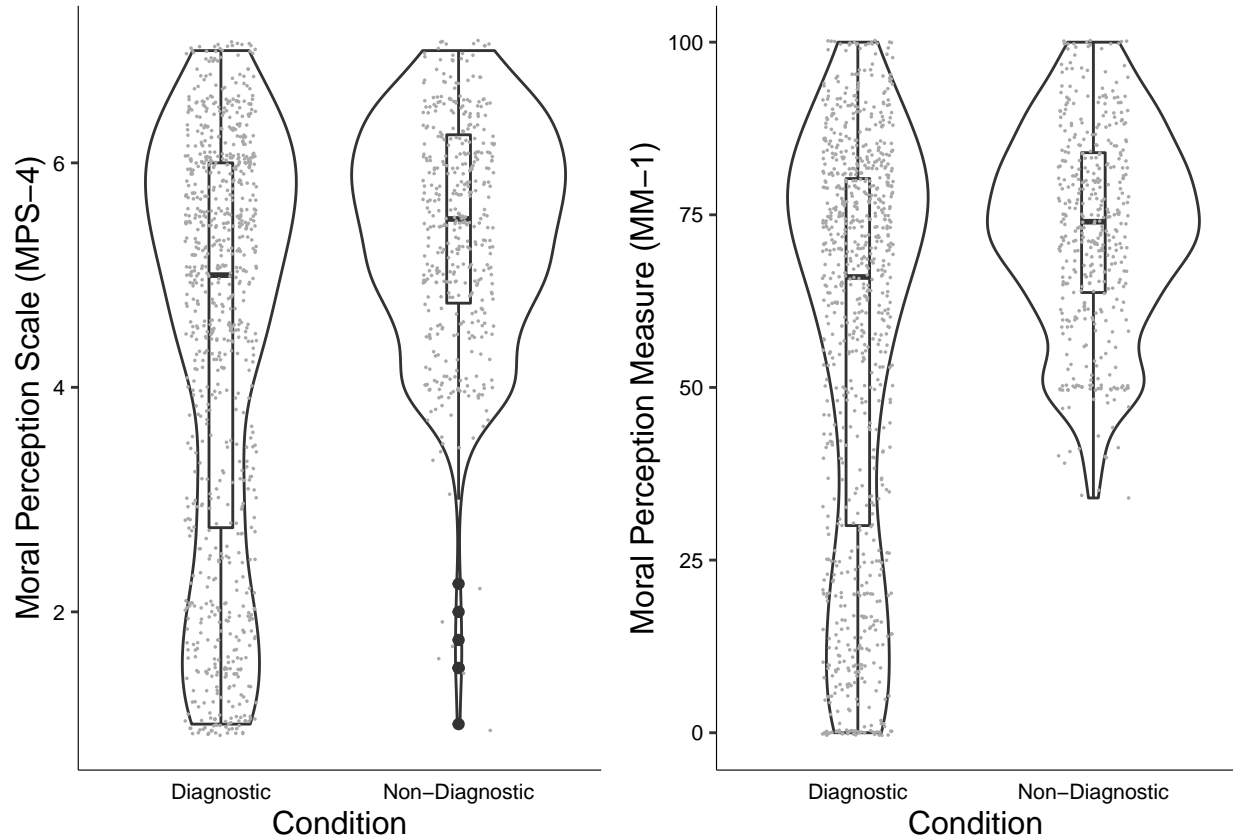
Figure 1

Pilot Study 1: Differences in moral perception depending on condition

responses. Our outcome measure was MM-1, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(2) = 924.82$, $p < .001$. Condition was a significant predictor in the model $b = $ -8.22, $t(210.98) = $ -8.60, $p < .001$, with the non-diagnostic (neutral) descriptions being rated as more moral than the diagnostic (morally relevant) descriptions, see Figure 1.

Study 1

Study 1: Method

The aim of Study 1 is to test if the dilution effect exists in the moral domain. Participants were presented with descriptions of four characters, two descriptions will only contain diagnostic information (morally relevant information) and two will additionally

contain non-diagnostic information (non morally relevant information) along with the diagnostic information. We hypothesize that moral perceptions of the diagnostic only descriptions will be more severe than for the descriptions that also contain non-diagnostic information.

Study 1: Participants and design

Study 1 was a within-subjects design. The independent variable was condition with two levels, diagnostic information only (diagnostic), and non-diagnostic information additionally included (non-diagnostic). We used the same two dependent variables as in Pilot Study 1, the four item moral perception scale (MPS-4) which showed good reliability, α = 0.83, and the single item moral perception measure MM-1.

A total sample of 901 (302 female, 523 male, 0 non-binary, 5 other; 2 prefer not to say, M_{age} = 26.16, min = 18, max = 76, SD = 10.14) started the survey. Participants were recruited from the student population at University of [BLINDED].

Participants who failed both manipulation checks were removed (n = 100), leaving a total sample of 801 participants (283 female, 496 male, 5 other, 5 prefer not to say; M_{age} = 26.25, min = 18, max = 76, SD = 10.20).

Study 1: Procedure and materials

As in the pilot study, data were collected using an online questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented with four descriptions of characters (*Sam, Alex, Francis, Robin* from Pilot Study 1). All descriptions included diagnostic information relating to three moral foundations, e.g., *Imagine a person named Robin. Throughout their life they have been known to physically hurt others, treat some people differently to others, and show lack of loyalty.* We programmed our survey to randomly present non-diagnostic information along with two of the descriptions participants read (this was done through blocking, for details on the blocks see full materials at https://osf.io/mdnpv/?view_only=77883e3fbc3d45f1a35fe92d5318cb67. Study 1 was

pre-registered at https://aspredicted.org/DVY_QN3

Study 1: Results

The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*, $M_{\text{MPS-4}} = 2.55$, $SD_{\text{MPS-4}} = 0.86$, *Francis*, $M_{\text{MPS-4}} = 3.05$, $SD_{\text{MPS-4}} = 0.97$, *Alex*, $M_{\text{MPS-4}} = 2.32$, $SD_{\text{MPS-4}} = 0.88$, *Robin*, $M_{\text{MPS-4}} = 2.13$, $SD_{\text{MPS-4}} = 0.91$. There was significant variation depending on the description, $F(2{,}279.86, 2.85) = 297.82$, $p < .001$, partial $\eta^2 = 0.13$. *Francis* appeared to be rated as more moral than each of the other characters (all $ps < .001$), while *Robin* was rated as less moral than each of the other characters (all $ps < .001$), while *Sam* was rated more favorably than *Alex* ($p < .001$).

The means and standard deviations for MM-1 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MM-1}} = 23.94$, $SD_{\text{MM-1}} = 16.18$; *Francis* (diagnostic/moral), $M_{\text{MM-1}} = 30.12$, $SD_{\text{MM-1}} = 17.86$; *Alex* (diagnostic/moral), $M_{\text{MM-1}} = 20.55$, $SD_{\text{MM-1}} = 16.65$; *Robin* (diagnostic/moral), $M_{\text{MM-1}} = 20.60$, $SD_{\text{MM-1}} = 17.06$. There was significant variation depending on the description, $F(2{,}253.09, 2.82) = 154.08$, $p < .001$, partial $\eta^2 = 0.05$. *Francis* was rated more favorably than all other characters ($p < .001$), *Sam* was the next most favorably rated character, rated significantly more favorably than both *Alex* and *Robin* ($ps < .001$), there was no difference between *Alex* and *Robin* ($p = = .953$).

We conducted a linear-mixed-effects model to test if condition influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants, and scenario was also included in the model. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(8) = 816.91$, $p < .001$. Condition significantly influenced responses to the MPS-4, $F(1, 799.42) = 51.47$, $p < .001$; and was a significant predictor in the model when controlling for scenario, $b = $ -0.09, $t(2{,}599.11) = $ -3.56, $p < .001$, with the non-diagnostic descriptions being rated as more moral than the diagnostic (morally relevant) descriptions of immoral characters Figure 2.
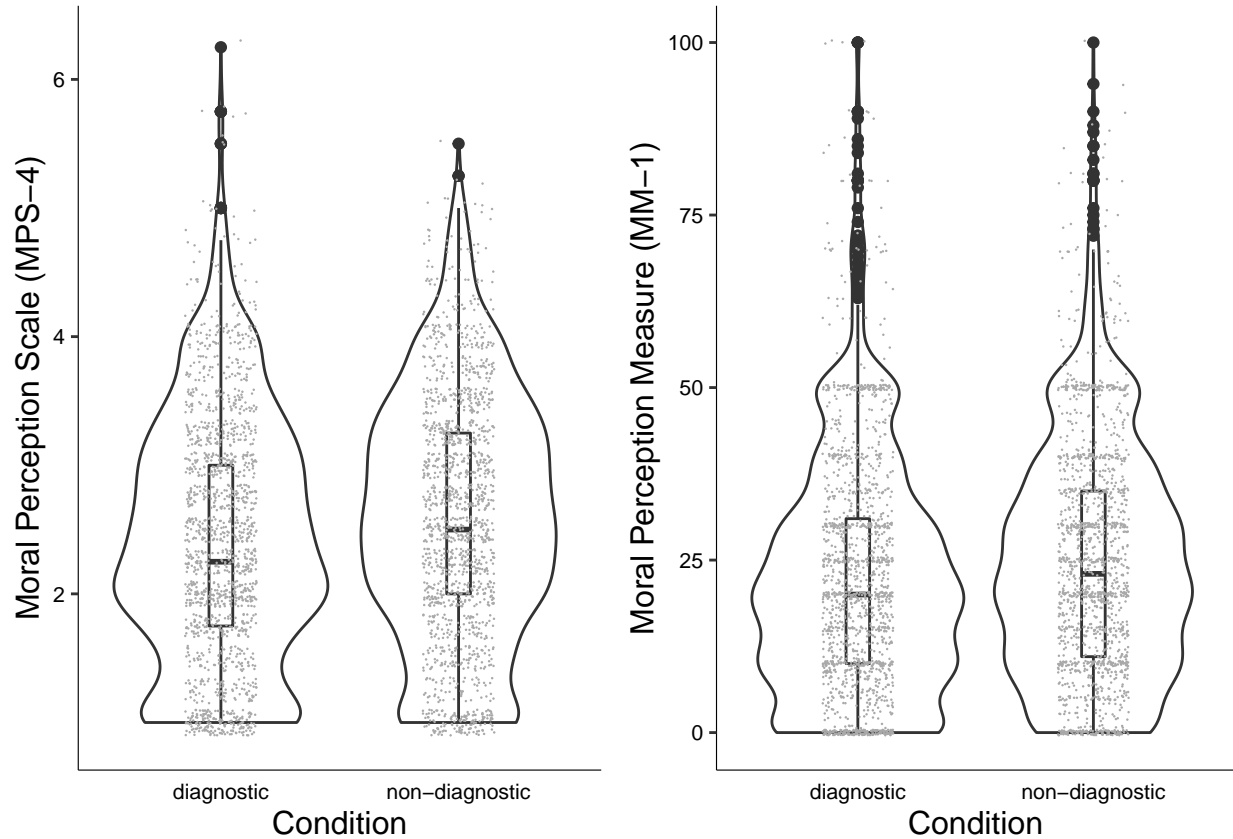
Figure 2

Study 1: Differences in moral perception depending on condition

We conducted a linear-mixed-effects model to test if condition influenced MM-1 responses. Our outcome measure was MM-1, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(8) = 475.52$, $p < .001$. Condition significantly predicted MM-1 responses $F(1, 799.71) = 44.39$, $p < .001$, and when controlling for scenario was a significant predictor in the model $b = -1.05$, $t(2,444.33) = -2.56$, $p = .010$, with the non-diagnostic (neutral) descriptions being rated as more moral than the diagnostic (morally relevant) descriptions, see Figure 2.

In the supplementary analyses we report the effect of condition on moral perception for each description individually.

Pilot Study 2

Study 1 showed the moral dilution effect for judgments of *bad* characters. The aim of this pilot study was to develop and test materials that may be used to study the moral dilution effect for judgments of morally *good* characters. As with Pilot Study 1, we developed diagnostic and non-diagnostic descriptions. We hypothesized that evaluations of the diagnostic descriptions would be more extreme (more moral) than for the non-diagnostic descriptions

Pilot Study 2: Method

Pilot 2: Participants and design

The pilot study was a within-subjects design. The independent variable was description type with two levels, *diagnostic* and *non-diagnostic*. We used the same two dependent variables as in previous studies, the four item moral perception scale (MPS-4, $\alpha = 0.84$), and the single item moral perception measure (MM-1).

A total sample of 245 (70 female, 175 male, 0 non-binary, 0 prefer not to say; $M_{\text{age}} = 36.69$, min = 18, max = 71, $SD = 9.57$) started the survey. Participants were recruited from MTurk.

We removed participants who failed both manipulation checks ($n = 30$), leaving a total sample of 215 participants (63 female, 152 male, 0 non-binary, 0 prefer not to say; $M_{\text{age}} = 36.59$, min = 18, max = 71, $SD = 9.59$).

Pilot 2: Procedure and materials

Data were collected using an online questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented with descriptions of six characters.

Moral character descriptions were developed by combining descriptions relating to three different moral foundations, focusing on upholding the moral foundations (rather than transgressions as in previous studies). A sample description reads: *Imagine a person named*

Sam. Throughout their life they have been known to always help and care for others, treat everyone fairly and equally, and show a strong sense of loyalty to others.. Full text of these descriptions can be found in the supplementary materials.

We developed neutral descriptions that included information relating to physical appearance/attributes, hobbies/activities, and a color preference, e.g., *Imagine a person named Charlie. They have blue eyes, drink coffee in the morning, and their favourite colour is green.*

We used the same gender ambiguous names, and we did not specify the gender of the characters. Pilot Study 2 was pre-registered at https://aspredicted.org/W52_VPX.

Pilot 2: Results

The means and standard deviations for MPS-4 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MPS-4}} = 6.01$, $SD_{\text{MPS-4}} = 0.91$, *Francis* (diagnostic/moral), $M_{\text{MPS-4}} = 5.89$, $SD_{\text{MPS-4}} = 0.95$, *Alex* (diagnostic/moral), $M_{\text{MPS-4}} = 5.94$, $SD_{\text{MPS-4}} = 0.94$, *Robin* (diagnostic/moral), $M_{\text{MPS-4}} = 5.93$, $SD_{\text{MPS-4}} = 0.92$, *Jackie* (non-diagnostic/neutral), $M_{\text{MPS-4}} = 5.60$, $SD_{\text{MPS-4}} = 0.99$, *Charlie* (non-diagnostic/neutral), $M_{\text{MPS-4}} = 5.53$, $SD_{\text{MPS-4}} = 1.08$. For the diagnostic descriptions, there was significant variation depending on the description, $F(613.27, 2.87) = 2.91$, $p = .036$, partial $\eta^2 = 0.00$, *Sam* was viewed significantly more favorably than *Francis* ($p = .040$). For the non-diagnostic descriptions there was no significant difference in ratings depending on description, $t(214) = -1.79$, $p = .075$, $d = 0.12$.

The means and standard deviations for MM-1 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MM-1}} = 79.85$, $SD_{\text{MM-1}} = 15.44$; *Francis* (diagnostic/moral), $M_{\text{MM-1}} = 78.30$, $SD_{\text{MM-1}} = 15.84$; *Alex* (diagnostic/moral), $M_{\text{MM-1}} = 79.78$, $SD_{\text{MM-1}} = 15.71$; *Robin* (diagnostic/moral), $M_{\text{MM-1}} = 79.46$, $SD_{\text{MM-1}} = 15.41$; *Jackie* (non-diagnostic/neutral), $M_{\text{MM-1}} = 73.44$, $SD_{\text{MM-1}} = 15.83$; *Charlie* (non-diagnostic/neutral), $M_{\text{MM-1}} = 73.07$, $SD_{\text{MM-1}} = 16.22$. For the diagnostic descriptions, we observed no significant variation depending on

the description, $F(594.47, 2.78) = 1.45$, $p = .231$, partial $\eta^2 = 0.002$. For the non-diagnostic descriptions there was no significant difference in ratings depending on description, $t(214) = -0.60$, $p = .552$, $d = 0.04$.

We conducted a linear-mixed-effects model to test if condition influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(2) = 475.42$, $p < .001$. Condition was a significant predictor in the model $b = 0.19$, $t(214.35) = 6.53$, $p < .001$, with the diagnostic descriptions being rated as more moral than the non-diagnostic descriptions of immoral characters Figure 3.
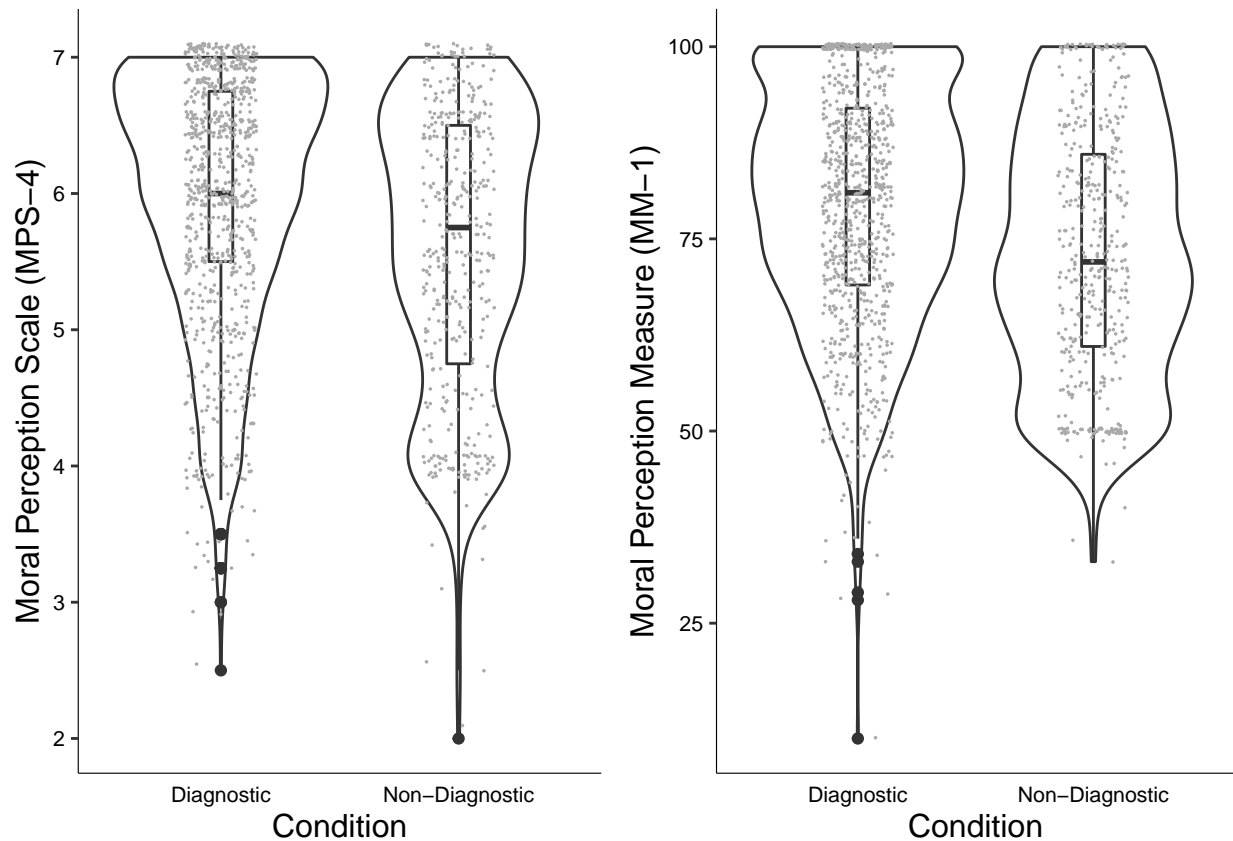

Figure 3

Pilot Study 2: Differences in moral perception depending on condition

We conducted a linear-mixed-effects model to test if condition influenced MM-1

responses. Our outcome measure was MM-1, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(2) = 324.13$, $p < .001$. Condition was a significant predictor in the model $b = 3.04$, $t(214.90) = 6.02$, $p < .001$, with the diagnostic descriptions being rated as more moral than the non-diagnostic descriptions, see Figure 3.

Study 2

Study 2: Method

The aim of Study 2 is to test if the dilution effect exists in the moral domain for judgments of morally *good* characters. Participants were presented with descriptions of four characters, two descriptions contain diagnostic information only (morally relevant information) and two will additionally contain non-diagnostic information (non morally relevant information) along with the diagnostic information. We hypothesize that moral perceptions of the diagnostic only descriptions will be more extreme (more moral) than for the descriptions that also contain non-diagnostic information.

Study 2: Participants and design

Study 2 was a within-subjects design. The independent variable was condition with two levels, diagnostic and non-diagnostic. We used the same two dependent variables as in previous studies, the four item moral perception scale (MPS-4, $\alpha = 0.85$), and the single item moral perception measure MM-1.

A total sample of 1068 (417.50 female, 555 male, 0 non-binary, 2 other; 3.17 prefer not to say, $M_{\text{age}} = 29.04$, min = 18, max = 74, $SD = 10.66$) started the survey. Participants were recruited from the student population at University of [BLINDED].

Participants who failed both manipulation checks were removed ($n = 248$), leaving a total sample of 820 participants (337 female, 466 male, 2 other, 2 prefer not to say; $M_{\text{age}} = 29.03$, min = 18, max = 74, $SD = 10.92$).

The majority of participants were from the student body: $n = 533$, (female = 370, male = 147, non-binary/other = 14, prefer not to say 3, $M_{age} = 25.50$, $SD = 9.60$).

In order to reach our pre-registered target sample size we recruited additional participants from MTurk: $n = 287$, (female = 96, male = 190, non-binary/other = 1, prefer not to say 1, $M_{age} = 35.70$, $SD = 10.10$).

Study 1: Procedure and materials

Again, data were collected using an online questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented with four descriptions of characters (*Sam, Alex, Francis, Robin* from Pilot Study 2). All descriptions included diagnostic information relating to three moral foundations, e.g., *Imagine a person named Alex. Throughout their life they have been known to protect and provide shelter to the weak and vulnerable, uphold the rights of others, and show respect for authority.* For each participant, two descriptions additionally included non-diagnostic information (this was randomized through blocking, see https://osf.io/mdnpv/?view_only=77883e3fbc3d45f1a35fe92d5318cb67. Study 1 was pre-registered at https://aspredicted.org/NX2_HN6

Study 2: Results

The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*, $M_{\text{MPS-4}} = 6.12$, $SD_{\text{MPS-4}} = 0.97$, *Francis*, $M_{\text{MPS-4}} = 5.86$, $SD_{\text{MPS-4}} = 1.07$, *Alex*, $M_{\text{MPS-4}} = 6.13$, $SD_{\text{MPS-4}} = 0.99$, *Robin*, $M_{\text{MPS-4}} = 6.10$, $SD_{\text{MPS-4}} = 0.99$. There was significant variation depending on the description, $F(2,355.68, 2.88) = 54.47$, $p < .001$, partial $\eta^2 = 0.01$. *Francis* appeared to be rated as less moral than each of the other characters (all $ps < .001$).

The means and standard deviations for MM-1 for each scenario are as follows: *Sam* (diagnostic/moral), $M_{\text{MM-1}} = 84.60$, $SD_{\text{MM-1}} = 14.47$; *Francis* (diagnostic/moral), $M_{\text{MM-1}} = 82.05$, $SD_{\text{MM-1}} = 15.24$; *Alex* (diagnostic/moral), $M_{\text{MM-1}} = 85.02$, $SD_{\text{MM-1}} = 15.01$; *Robin* (diagnostic/moral), $M_{\text{MM-1}} = 84.95$, $SD_{\text{MM-1}} = 13.94$. There was significant variation depending on the description, $F(2,386.54, 2.91) = 24.20$, $p < .001$, partial $\eta^2 = 0.007$.

Francis was rated less favorably than all other characters (all $ps < .001$).

We conducted a linear-mixed-effects model to test if condition influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was condition; we allowed intercepts and the effect of condition to vary across participants, and scenario was also included in the model. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(8) = 160.00$, $p < .001$. Condition did not influence responses to the MPS-4, $F(1, 838.11) = 0.24$, $p = .624$; and was not a significant predictor in the model when controlling for scenario, $b = 0.00$, $t(2,403.47) = 0.24$, $p = .814$, see Figure 4.
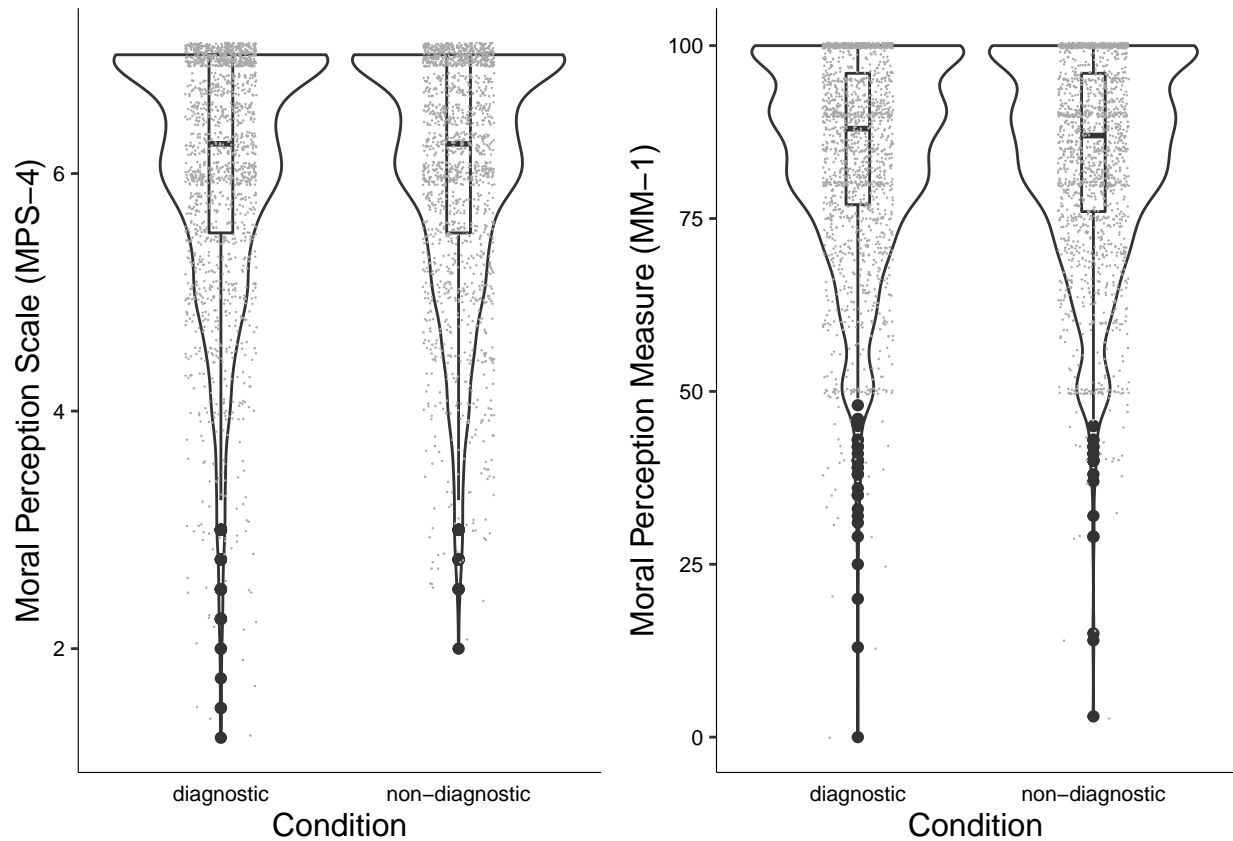

Figure 4

Study 2: Differences in moral perception depending on condition

We conducted a linear-mixed-effects model to test if condition influenced MM-1 responses. Our outcome measure was MM-1, our predictor variable was condition; we

allowed intercepts and the effect of condition to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(8) = 75.69$, $p < .001$. Condition did not influence MM-1 responses $F(1, 2{,}453.06) = 1.23$, $p = .267$, and was not a significant predictor in the model $b = -0.30$, $t(2{,}654.99) = -0.90$, $p = .366$, see Figure 4.

In the supplementary analyses we report the effect of condition on moral perception for each description individually.

Discussion

Accessibility Statement

All data and analysis code are publicly available on this project's OSF page at

https://osf.io/mdnpv/?view_only=77883e3fbc3d45f1a35fe92d5318cb67.

References

Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., . . . Eden, A. (2020). Validating the extended character morality questionnaire. *Media Psychology*, *23*(1), 107–130. https://doi.org/10.1080/15213269.2019.1572523

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2022). Moral Judgment as Categorization (MJAC). *Perspectives on Psychological Science*, *17*(1), 131–152. https://doi.org/10.1177/1745691621990636

Schein, C., & Gray, K. J. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Walker, A. C., Turpin, M. H., Fugelsang, J. A., & Białek, M. (2021). Better the two devils you know, than the one you don't: Predictability influences moral judgments of immoral actors. *Journal of Experimental Social Psychology*, *97*, 104220. https://doi.org/10.1016/j.jesp.2021.104220