Study 3

Blinded[1], Blinded[2], Blinded[1], and & Blinded[1]

[1] Blinded

[2] Blinded

Author Note

Correspondence concerning this article should be addressed to Blinded, Blinded.

E-mail: Blinded

Abstract

Six studies etc.

Keywords: keywords

Word count: TBC

Study 3

Study 3 - Good and Bad Characters

In Study 1 we found evidence for the moral dilution effect for judgments of *bad* moral characters. In Study 2 we failed replicate this effect for judgments of *good* moral characters. The aim of Study 3 was to test if valence (good vs. bad) moderates the moral dilution effect. We hypothesized that valence (good vs bad) would interact with condition in producing a dilution effect, such that the dilution effect would be observed for bad characters but not for good characters. Study 3 was pre-registered at https://aspredicted.org/QDF_XT1.

Study 3: Method

Study 3: Participants and design

Study 3 was a 2 × 2 within-subjects factorial design. The first independent variable was condition with two levels, diagnostic and non-diagnostic. The second independent variable was valence of character description, with two levels morally good and morally bad. We used the same two dependent variables as in previous studies, the four item moral perception scale (MPS-4, $\alpha = 0.97$), and the single item moral perception measure MM-1.

A total sample of 1386 (535 female, 758 male, 10 non-binary, 2 other; 11 prefer not to say, $M_{\text{age}} = 29.67$, min = 0.36, max = 70, $SD = 8.97$) started the survey. Participants were recruited from Prolific Academic and paid $0.40 for their participation.

Participants who failed both manipulation checks ($n = 541$), or did not complete all measures were removed , leaving a total sample of 814 participants (462 female, 327 male, 2 other, 2 prefer not to say; $M_{\text{age}} = 26.03$, min = 11, max = 70, $SD = 9.53$).

Study 3: 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 taken from Studies 1 and 2. To ensure consistency across character judgments, we selected descriptions that related to the same moral foundations (care, fairness, and loyalty). We used the same four character names as in previous studies. The *good* characters were *Sam* and *Robin*, and the *bad* characters were *Francis* and *Alex*, e.g., *Imagine a person named Robin. Throughout their life they have been known to show compassion and empathy for others, act with a sense of fairness and justice, and, never to break their word.* or, *Imagine a person named Alex. Throughout their life they have been known to be cruel, act unfairly, and to betray their own group.* Full descriptions for each character are in the supplementary materials. One description for each the *good* and *bad* characters was randomly assigned to include non-diagnostic information for each participant thus all participants were exposed to all conditions (see https://osf.io/mdnpv/?view_only=77883e3fbc3d45f1a35fe92d5318cb67for details of the randomization blocks). Study 3 was pre-registered at https://aspredicted.org/QDF_XT1

Study 3: Results

As in Studies 1 and 2 we assessed the quality of the data by examining the extent to which responses fell above / below the midpoint for the *bad* / *good* descriptions respectively. All $N = 814$ participants responded to two *bad* descriptions and 2 *good* descriptions, resulting in a total of 1628 responses for each measure for both *bad* and *good* descriptions. Taking the *bad* descriptions first, for MPS-4, 33 (2.03%) responses were above the midpoint, and for MM-1, 114 (7.00%) were above the midpoint.

Regarding the *good* descriptions, for MPS-4, 22 (1.35%) responses were below the midpoint, and for MM-1, 17 (1.04%) were below the midpoint.

In order to test for the information type $\times$ valence interaction effect in a single

model recoded both MPS-4 and MM-1 into two new variables MPS-4R, and MM-1R. These recoded variables were the same as the original variables but the responses to the good characters were reverse coded. This allowed us to examine whether the dilution effect was different depending on whether participants were judging good characters or bad characters without this analysis being confounded by valence.

First we conducted a within subjects factorial ANOVA to test for differences in responses to MPS-4R depending on information type and valence. There was a main effect for condition, $F(1, 813) = 48.53$, $p < .001$, partial $\eta^2 = 0.06$, 95% CI [0.03, 0.09], and a main effect for valence, $F(1, 813) = 3{,}383.29$, $p < .001$, partial $\eta^2 = 0.81$, 95% CI [0.79, 0.82]. There was no condition × valence interaction effect, $F(1, 813) = 6.35$, $p = .012$, partial $\eta^2 = 0.01$, 95% CI [0.00, 0.02].

Follow-up pairwise t-tests indicated that for the *bad* characters there were significant differences in MPS-4 responses depending on information type $t(813)$, = 6.59, p = < .001 (p_{adjusted} = < .001), $d = 0.23$, 95% CI [0.13, 0.25]. MPS-4 responses were higher in the non-diagnostic condition ($M = 2.33$, $SD = 0.86$) compared to the diagnostic condition ($M = 2.14$, $SD = 0.81$).

Similarly, for *good* characters, were significant differences in MPS-4 responses depending on information type $t(813)$, = -3.43, p = < .001 (p_{adjusted} = = .003), $d = 0.12$, 95% CI [-0.15, -0.04]. MPS-4 responses were lower in the non-diagnostic condition (M = 6.21, $SD = 0.84$) compared to the diagnostic condition ($M = 6.31$, $SD = 0.74$).

Next we conducted a within subjects factorial ANOVA to test for differences in responses to MM-1R depending on information type and valence. There was a main effect for condition, $F(1, 813) = 48.53$, $p < .001$, partial $\eta^2 = 0.04$, 95% CI [0.01, 0.06], and a main effect for valence, $F(1, 813) = 3{,}383.29$, $p < .001$, partial $\eta^2 = 0.24$, 95% CI [0.19, 0.29]. There was no condition × valence interaction effect, $F(1, 813) = 6.35$, $p = .012$, partial $\eta^2 = 0.00$, 95% CI [0, 0.01].

Follow-up pairwise t-tests indicated that for the *bad* characters there were significant differences in MM-1 responses depending on information type $t(813)$, = 4.27, p = < .001 (p_{adjusted} = < .001), d = 0.15, 95% CI [1.22, 3.29]. MPS-4 responses were higher in the non-diagnostic condition (M = 22.60, SD = 17.06) compared to the diagnostic condition (M = 20.35, SD = 17.13).

Similarly, for *good* characters, were significant differences in MPS-4 responses depending on information type $t(813)$, = -3.60, p = < .001 (p_{adjusted} = = .001), d = 0.13, 95% CI [-2.16, -0.64]. MPS-4 responses were lower in the non-diagnostic condition (M = 86.68, SD = 13.86) compared to the diagnostic condition (M = 88.08, SD = 11.94).

We conducted a linear-mixed-effects model to test if our predictors influenced MPS-4R responses. Our outcome measure was MPS-4R, our predictor variables were condition and valence; we allowed intercepts and the effects of condition and valence to vary across participants, and we included random effects for scenario. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(5)$ = 2,695.60, p < .001. Overall, there was a significant main effect for condition, $F(1, 813)$ = 48.53, p < .001; valence significantly predicted responses, $F(1, 813)$ = 3,383.29, p < .001; and there was a significant condition × valence interaction, $F(1, 813)$ = 6.35, p = .012.

We conducted a linear-mixed-effects model to test if our predictors influenced MM-1R responses. The model was the same as the previous model, with a change to the outcome measure, our outcome measure for this model was MM-1R. As above, our predictor variables were condition and valence; we allowed intercepts and the effects of condition and valence to vary across participants, and we included random effects for scenario. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model, $\chi^2(5)$ = 749.17, p < .001. Overall there was a main effect for condition, $F(1, 813)$ = 29.92, p < .001; valence significantly predicted

responses, $F(1, 813) = 258.78$, $p < .001$; and there was no significant condition × valence interaction, $F(1, 813) = 1.78$, $p = .183$.

Interestingly, there was a consistent effect for both condition and valence, as well as a condition × valence interaction effect. To further examine this interaction effect, we report separate analyses for the good and bad descriptions below.

Differences in the *Bad* Descriptions

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, $\chi^2(3) = 76.88$, $p < .001$. Condition significantly influenced MPS-4 responses $F(1, 812.00) = 46.02$, $p < .001$, and was a significant predictor in the model $b =$ -0.10 ($b_{\text{standardized}} =$ -0.11), $t(812.00) =$ -6.78, $p < .001$, ($d =$ -0.23), see Figure 1.

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, $\chi^2(3) = 46.32$, $p < .001$. Condition significantly influenced MM-1 responses $F(1, 812.00) = 19.25$, $p < .001$, and was a significant predictor in the model $b =$ -1.14 ($b_{\text{standardized}} =$ -0.06), $t(812.00) =$ -4.39, $p < .001$, ($d =$ -0.15), see Figure 1.

Differences in the *Good* Descriptions

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, $\chi^2(3) = 30.01$, $p < .001$. Condition significantly influenced MPS-4 responses $F(1, 812.00) = 11.87$, $p < .001$, and was a significant predictor in the model $b = 0.05$ ($b_{\text{standardized}} = 0.06$), $t(812.00) = 3.45$, $p < .001$, ($d = 0.12$), see Figure 1.

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, $\chi^2(3) = 41.07$, $p < .001$. Condition significantly influenced MM-1 responses $F(1, 812) = 13.21$, $p < .001$, and was a significant predictor in the model $b = 0.69$ ($b_{\text{standardized}} = 0.05$), $t(812) = 3.63$, $p < .001$, ($d = 0.13$), see Figure 1.

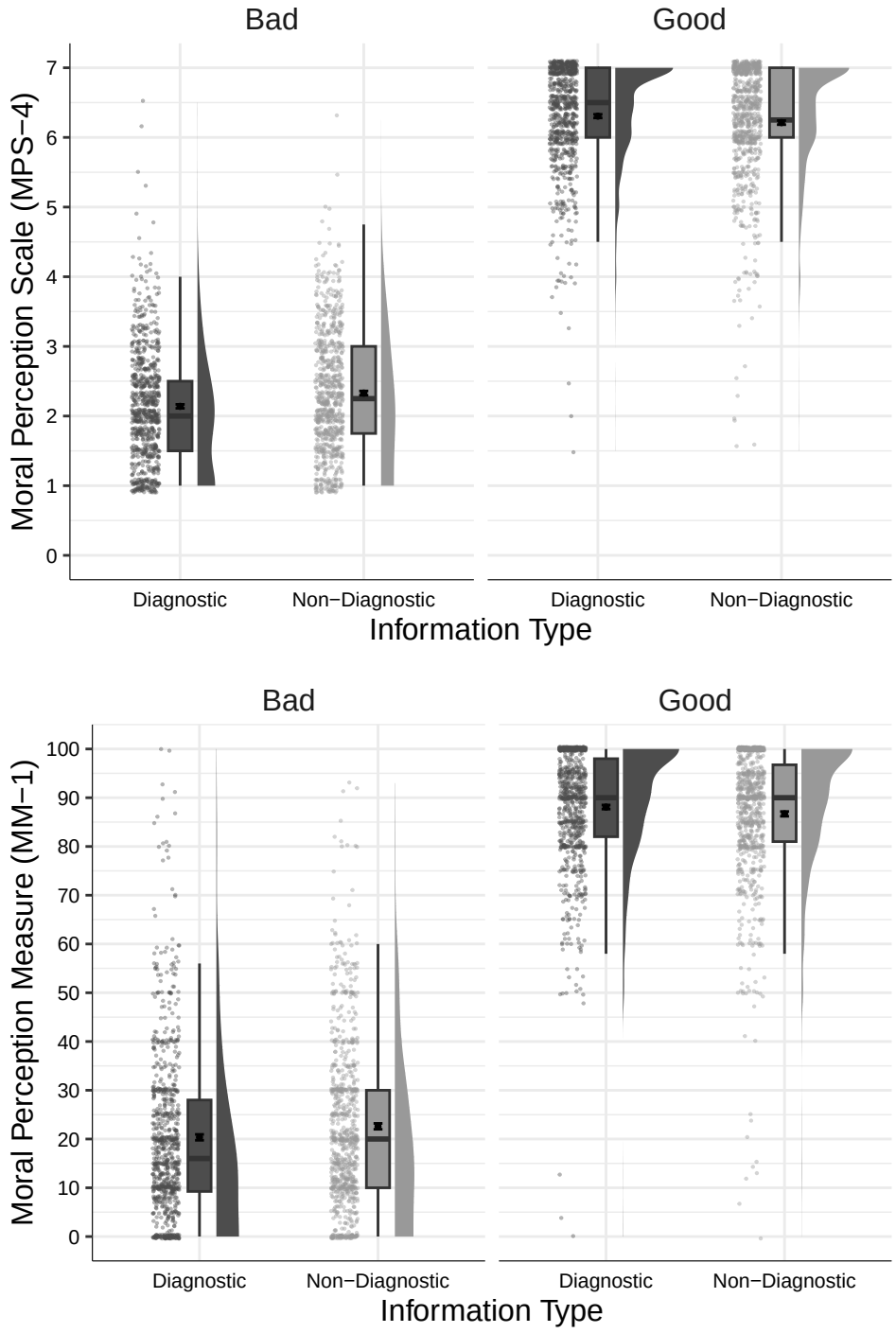

Figure 1

Study 3: Differences in moral perception depending on condition