

### **Study 3**

Blinded<sup>1</sup>, Blinded<sup>2</sup>, Blinded<sup>1</sup>, and & Blinded<sup>1</sup>

<sup>1</sup> Blinded

<sup>2</sup> 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](https://aspredicted.org/QDF_XT1).

### Study 3: Method

#### *Study 3: Participants and design*

Study 3 was a  $2 \times 2$  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 (758 female, 535 male, 10 non-binary, 2 other; 11 prefer not to say,  $M_{\text{age}} = 29.67$ ,  $\text{min} = 0.36$ ,  $\text{max} = 70$ ,  $SD = 8.97$ ) started the survey. Participants were recruited from MTurk and paid \$0.40 for their participation.

Participants who failed both manipulation checks were removed ( $n = 541$ ), leaving a total sample of 828.75 participants (327 female, 462 male, 2 other, 2 prefer not to say;  $M_{\text{age}} = 26.03$ ,  $\text{min} = 11$ ,  $\text{max} = 70$ ,  $SD = 9.53$ ).

### ***Study 3: Procedure and materials***

Again, data were collected using an online questionnaire presented with Qualtrics ([www.qualtrics.com](http://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=77883e3fbc3d45f1a35fe92d5318cb67](https://osf.io/mdnpv/?view_only=77883e3fbc3d45f1a35fe92d5318cb67) for details of the randomization blocks). Study 3 was pre-registered at [https://aspredicted.org/QDF\\_XT1](https://aspredicted.org/QDF_XT1)

### **Study 3: Results**

The means and standard deviations for MPS-4 for each scenario are as follows: *Sam* (good),  $M_{\text{MPS-4}} = 6.20$ ,  $SD_{\text{MPS-4}} = 0.81$ , *Francis* (bad),  $M_{\text{MPS-4}} = 2.15$ ,  $SD_{\text{MPS-4}} = 0.79$ , *Alex* (bad),  $M_{\text{MPS-4}} = 2.32$ ,  $SD_{\text{MPS-4}} = 0.89$ , *Robin* (good),  $M_{\text{MPS-4}} = 6.32$ ,  $SD_{\text{MPS-4}} = 0.76$ . There was significant variation depending on the description,  $F(2,1515) = 6,251.52$ ,  $p < .001$ , partial  $\eta^2 = 0.86$ . Both the *good* characters (*Robin* and *Sam*) were rated significantly more favorably than both the *bad* characters (*Alex* and *Francis*; all  $ps < .001$ ). There were no differences between *Robin* and *Sam* (good:  $p = .003$ ) or between *Alex* and *Francis* (bad;  $p < .001$ ).

The means and standard deviations for MM-1 for each scenario are as follows: *Sam*

(good),  $M_{\text{MM-1}} = 86.33$ ,  $SD_{\text{MM-1}} = 13.68$ ; *Francis* (bad),  $M_{\text{MM-1}} = 20.25$ ,  $SD_{\text{MM-1}} = 16.93$ ; *Alex* (bad),  $M_{\text{MM-1}} = 23.07$ ,  $SD_{\text{MM-1}} = 17.51$ ; *Robin* (good),  $M_{\text{MM-1}} = 88.44$ ,  $SD_{\text{MM-1}} = 12.00$ . There was significant variation depending on the description,  $F(2,1380) = 5,282.47$ ,  $p < .001$ , partial  $\eta^2 = 0.826$ . Again, the *good* characters (*Robin* and *Sam*) were rated significantly more favorably than the *bad* characters (*Alex* and *Francis*; all  $ps < .001$ ). There were no differences between *Robin* and *Sam* (*good*:  $p = .001$ ) or between *Alex* and *Francis* (*bad*;  $p < .001$ ).

We conducted a linear-mixed-effects model to test if our predictors influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variables were condition and valence; we allowed intercepts and the effects of condition and valence to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model,  $\chi^2(5) = 2,733.13$ ,  $p < .001$ . Overall, there was a significant main effect for condition,  $F(1, 825) = 47.54$ ,  $p < .001$ ; valence significantly predicted responses,  $F(1, 823) = 3,423.33$ ,  $p < .001$ ; and there was a significant condition  $\times$  valence interaction,  $F(1, 819) = 6.56$ ,  $p = .011$ .

We conducted a linear-mixed-effects model to test if our predictors influenced MM-1 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-1. As above, our predictor variables were condition and valence; we allowed intercepts and the effects of condition and valence to vary across participants. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model,  $\chi^2(5) = 6,652.62$ ,  $p < .001$ . Overall there was a main effect for condition,  $F(1, 829) = 30.29$ ,  $p < .001$ ; valence significantly predicted responses,  $F(1, 824) = 273.59$ ,  $p < .001$ ; and there was no significant condition  $\times$  valence interaction,  $F(1, 825) = 1.57$ ,  $p = .211$ .

For both MP-4 and MM-1 we found a main effect for condition and valence, and there was a condition  $\times$  valence interaction for MP-4 but not for MM-1. We conducted

follow-up analyses to test the if the main effect for condition holds for both good and bad descriptions separately.

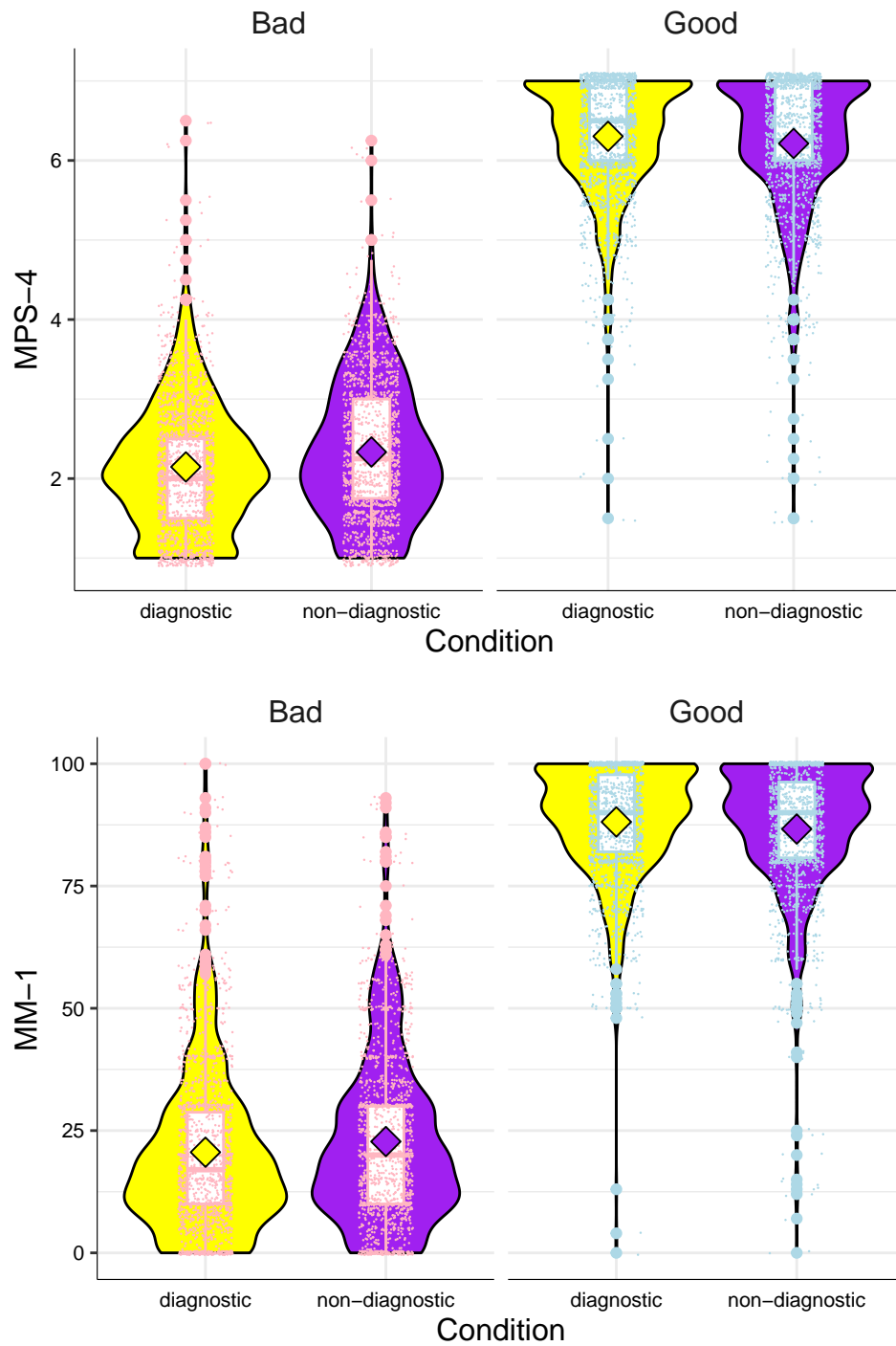
### Differences in the *Bad* Descriptions

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,  $\chi^2(3) = 77.91$ ,  $p < .001$ . Condition significantly influenced MPS-4 responses  $F(1, 818.61) = 46.18$ ,  $p < .001$ , and was a significant predictor in the model  $b = -0.10$ ,  $t(818.61) = -6.80$ ,  $p < .001$ , 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,  $\chi^2(3) = 47.56$ ,  $p < .001$ . Condition significantly influenced MM-1 responses  $F(1, 822.47) = 19.21$ ,  $p < .001$ , and was a significant predictor in the model  $b = -1.13$ ,  $t(822.47) = -4.38$ ,  $p < .001$ , 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,  $\chi^2(3) = 31.37$ ,  $p < .001$ . Condition significantly influenced MPS-4 responses  $F(1, 826.37) = 11.61$ ,  $p < .001$ , and was a significant predictor in the model  $b = 0.05$ ,  $t(826.37) = 3.41$ ,  $p < .001$ , see Figure 1.

**Figure 1**

*Study 3: 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,  $\chi^2(3) = 43.43$ ,  $p < .001$ . Condition significantly influenced MM-1 responses  $F(1, 825) = 13.96$ ,  $p < .001$ , and was a significant predictor in the model  $b = 0.71$ ,  $t(825) = 3.74$ ,  $p < .001$ , see Figure 1.