

¹ Supplement: Moral Dilution² Cillian McHugh¹ & Eric R. Igou¹³ ¹ University of Limerick⁴ Author Note

⁵ Department of Psychology, University of Limerick. All procedures performed in
⁶ studies involving human participants were approved by institutional research ethics
⁷ committee and conducted in accordance with the Code of Professional Ethics of the
⁸ Psychological Society of Ireland, and with the 1964 Helsinki declaration and its later
⁹ amendments or comparable ethical standards. Informed consent was obtained from all
¹⁰ individual participants included in the study. The authors declare that there are no
¹¹ potential conflicts of interest with respect to the research, authorship, and/or publication
¹² of this article. All authors consented to the submission of this manuscript.

¹³ Correspondence concerning this article should be addressed to Cillian McHugh,
¹⁴ University of Limerick, Limerick, Ireland, V94 T9PX. E-mail: cillian.mchugh@ul.ie

15

Abstract

- 16 Supplementary analysis to accompany the manuscript The Moral Dilution Effect:
17 Irrelevant Information Influences Judgments of Moral Character.

18 *Keywords:* keywords

19 Word count: X

20 Supplement: Moral Dilution

21 **Contents**

22 Materials	4
23 Study 1 (bad): Supplementary Analyses	8
24 Study 2 (good): Supplementary Analyses	12
25 Study 3 (bad and good): Supplementary Analyses	16
26 Pilot Study 1	21
27 Pilot Study 2	26
28 Study S1 - Good Characters	31
29 Study S2 - Good and Bad Characters	37
30 Study S3 - Good and Bad Characters	45
31 References	52

32

Materials

33 **Descriptions (Bad Characters)**

34 **Diagnostic Descriptions.** Each moral description contains descriptive information
35 relating to three different moral foundations as follows: *Sam*: care, fairness, loyalty; *Robin*:
36 care, fairness, loyalty; *Francis*: purity, authority, fairness; *Alex*: care, fairness, authority.

37 **Sam.** Imagine a person named Sam. Throughout their life they have been known to
38 be cruel, act unfairly, and to betray their own group.

39 **Robin.** Imagine a person named Robin. Throughout their life they have been
40 known to physically hurt others, treat some people differently to others, and show lack of
41 loyalty.

42 **Francis.** Imagine a person named Francis. Throughout their life they have been
43 known to violate the standards of purity and decency, show lack of respect for authority,
44 and treat people unequally.

45 **Alex.** Imagine a person named Alex. Throughout their life they have been known
46 to cause others to suffer emotionally, to deny others their rights, and to cause chaos or
47 disorder.

48 **Non-Diagnostic Descriptions.**

49 **Jackie.** Imagine a person named Jackie. They have red hair, play tennis four times
50 a month, and have one older sibling and one younger sibling.

51 **Charlie.** Imagine a person named Charlie. They are left-handed, drink tea in the
52 morning, and have two older siblings and one younger sibling.

53 **Descriptions (Good Characters)**

54 **Diagnostic Descriptions.** Each moral description contains descriptive information
55 relating to three different moral foundations as follows: *Sam*: care, fairness, loyalty; *Robin*:
56 care, fairness, loyalty; *Francis*: purity, authority, fairness; *Alex*: care, fairness, authority.

57 **Sam.** Imagine a person named Sam. Throughout their life they have been known to
58 always help and care for others, treat everyone fairly and equally, and show a strong sense
59 of loyalty to others.

60 **Robin.** Imagine a person named Robin. Throughout their life they have been
61 known to show compassion and empathy for others, act with a sense of fairness and justice,
62 and, never to break their word.

63 **Francis.** Imagine a person named Francis. Throughout their life they have been
64 known to uphold the standards of purity and decency, show respect for authority, and to
65 always act honestly and fairly.

66 **Alex.** Imagine a person named Alex. Throughout their life they have been known
67 to protect and provide shelter to the weak and vulnerable, uphold the rights of others, and
68 show respect for authority.

69 **Non-Diagnostic**

70 **Jackie.** Imagine a person named Jackie. They have dark hair, go for a jog twice a
71 week, and their favorite color is blue.

72 **Charlie.** Imagine a person named Charlie. They have blue eyes, drink coffee in the
73 morning, and their favorite color is green.

⁷⁴ **Descriptions (Both Good and Bad)**

⁷⁵ **Diagnostic Descriptions.**

⁷⁶ ***Sam (good).*** Imagine a person named Sam. Throughout their life they have been
⁷⁷ known to always help and care for others, treat everyone fairly and equally, and show a
⁷⁸ strong sense of loyalty to others.

⁷⁹ ***Robin (good).*** Imagine a person named Robin. Throughout their life they have
⁸⁰ been known to show compassion and empathy for others, act with a sense of fairness and
⁸¹ justice, and, never to break their word.

⁸² ***Alex (bad).*** Imagine a person named Alex. Throughout their life they have been
⁸³ known to be cruel, act unfairly, and to betray their own group.

⁸⁴ ***Francis (bad).*** Imagine a person named Francis. Throughout their life they have
⁸⁵ been known to physically hurt others, treat some people differently to others, and show
⁸⁶ lack of loyalty.

⁸⁷ **Non Diagnostic Descriptions.** They have red hair, play tennis four times a
⁸⁸ month, and have one older sibling and one younger sibling.

⁸⁹ They are left-handed, drink tea in the morning, and have two older siblings and one
⁹⁰ younger sibling.

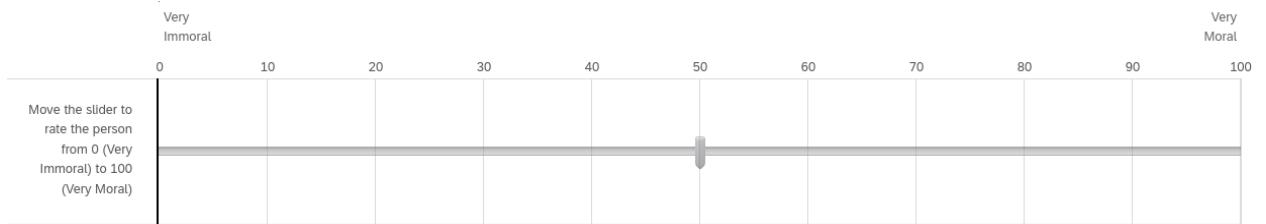
91 **Measures**

92 **Four-item Moral Perception Scale (MPS-4).** Please rate _____ along the
 93 following dimensions:

	1	2	3	4	5	6	7	
Bad	<input type="radio"/>	Good						
Immoral	<input type="radio"/>	Moral						
Violent	<input type="radio"/>	Peaceful						
Merciless	<input type="radio"/>	Empathetic						

Figure 1. Screenshot of the MPS-4 items as presented to participants

94 **Single-item Moral Perception Measure (MM-1).** Please rate _____
 95 according to immoral or moral you view them:



97

Study 1 (bad): Supplementary Analyses

98

Study 1: Combined Measure

99 We developed a combined moral perception measure by calculating the mean of the
100 combined mean-centered scores for MPS-4 and MM-1, and mean-centering this result.

101 Below we report the analyses for this combined measure.

102 The means and standard deviations for the combined measure for each scenario are
103 as follows: *Sam*, $M = 0.02$, $SD = 0.89$, *Francis*, $M = 0.48$, $SD = 1.00$, *Alex*, $M = -0.21$, SD
104 = 0.92, *Robin*, $M = -0.32$, $SD = 0.94$. There was significant variation depending on the
105 description, $F(3,2255) = 269.01$, $p < .001$, partial $\eta^2 = 0.10$. *Francis* appeared to be rated
106 as the most favorable, followed by *Sam*, then *Alex* and finally *Robin* as the least favorable
107 (all $ps < .001$).

108 We conducted a linear-mixed-effects model to test if condition influenced moral
109 perception. Our outcome measure was the combined moral perception measure, our
110 predictor variable was condition; we allowed intercepts and the effect of condition to vary
111 across participants, and scenario was also included in the model. Overall, the model
112 significantly predicted participants responses, and provided a better fit for the data than
113 the baseline model, $\chi^2(8) = 762.31$, $p < .001$. Condition significantly influenced responses
114 to the MPS-4, $F(1, 799.66) = 57.93$, $p < .001$; and was a significant predictor in the model
115 when controlling for scenario, $b = -0.08$, $t(2,501.32) = -3.42$, $p < .001$, with the
116 non-diagnostic descriptions being rated as more moral than the diagnostic (morally
117 relevant) descriptions of immoral characters Figure 2.

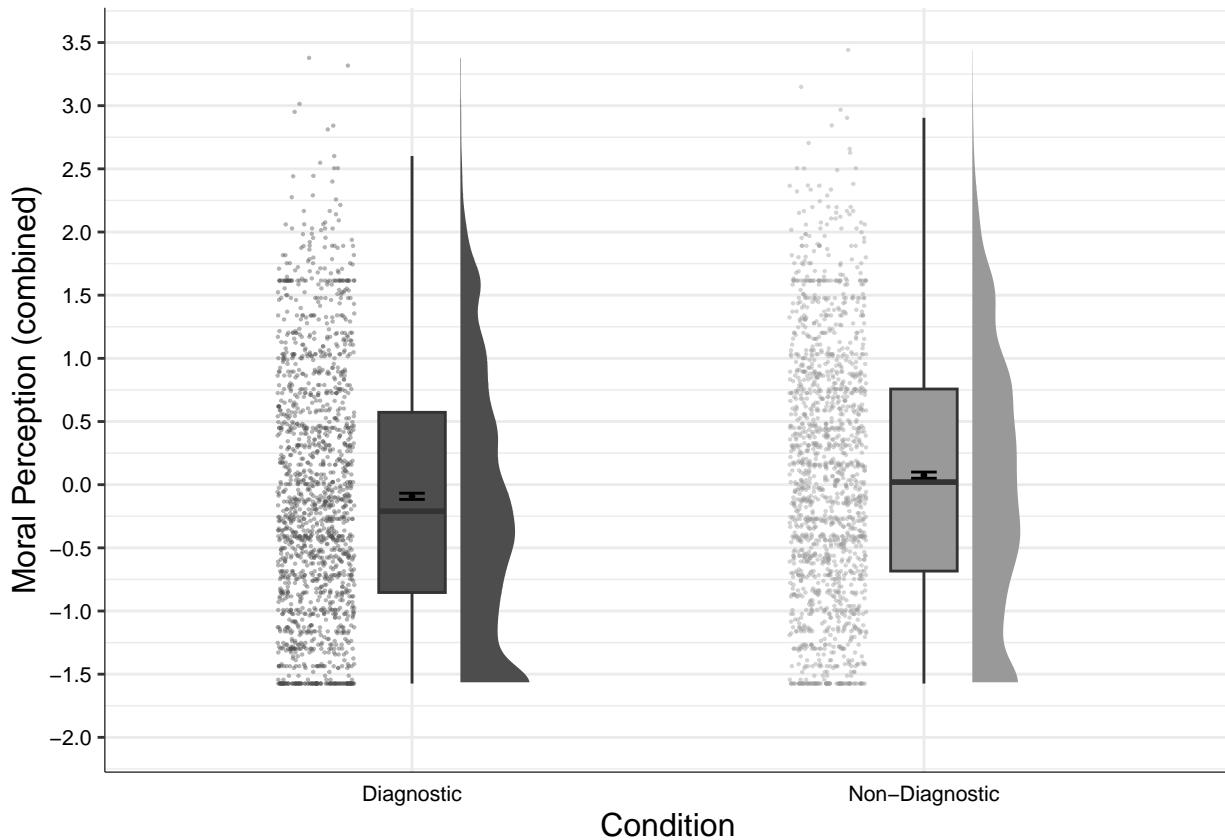


Figure 2. Study 1: Differences in combined measure depending on condition

¹¹⁸ **Study 1: Differences between the Descriptions**

¹¹⁹ We additionally conducted separate analyses for each scenario individually (for each
¹²⁰ dependent measure MPS-4, MM-1 and the combined measure). The responses for each
¹²¹ scenario across each measure depending on condition are displayed in Figure 3.

¹²² For *Sam*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
¹²³ = 2.70, SD = 0.82), than in the diagnostic condition (M = 2.42, SD = 0.87), $t(798.90)$ =
¹²⁴ -4.66, p < .001, d = 0.33; MM-1 ratings were higher in the non-diagnostic condition (M =
¹²⁵ 26.55, SD = 16.41), than in the diagnostic condition (M = 21.50, SD = 15.59), $t(787.84)$ =
¹²⁶ -4.45, p < .001, d = 0.32. For the combined measure ratings were also higher in the
¹²⁷ non-diagnostic condition (M = 0.18, SD = 0.88), than in the diagnostic condition (M =
¹²⁸ -0.13, SD = 0.88), $t(795.41)$ = -4.98, p < .001, d = 0.35.

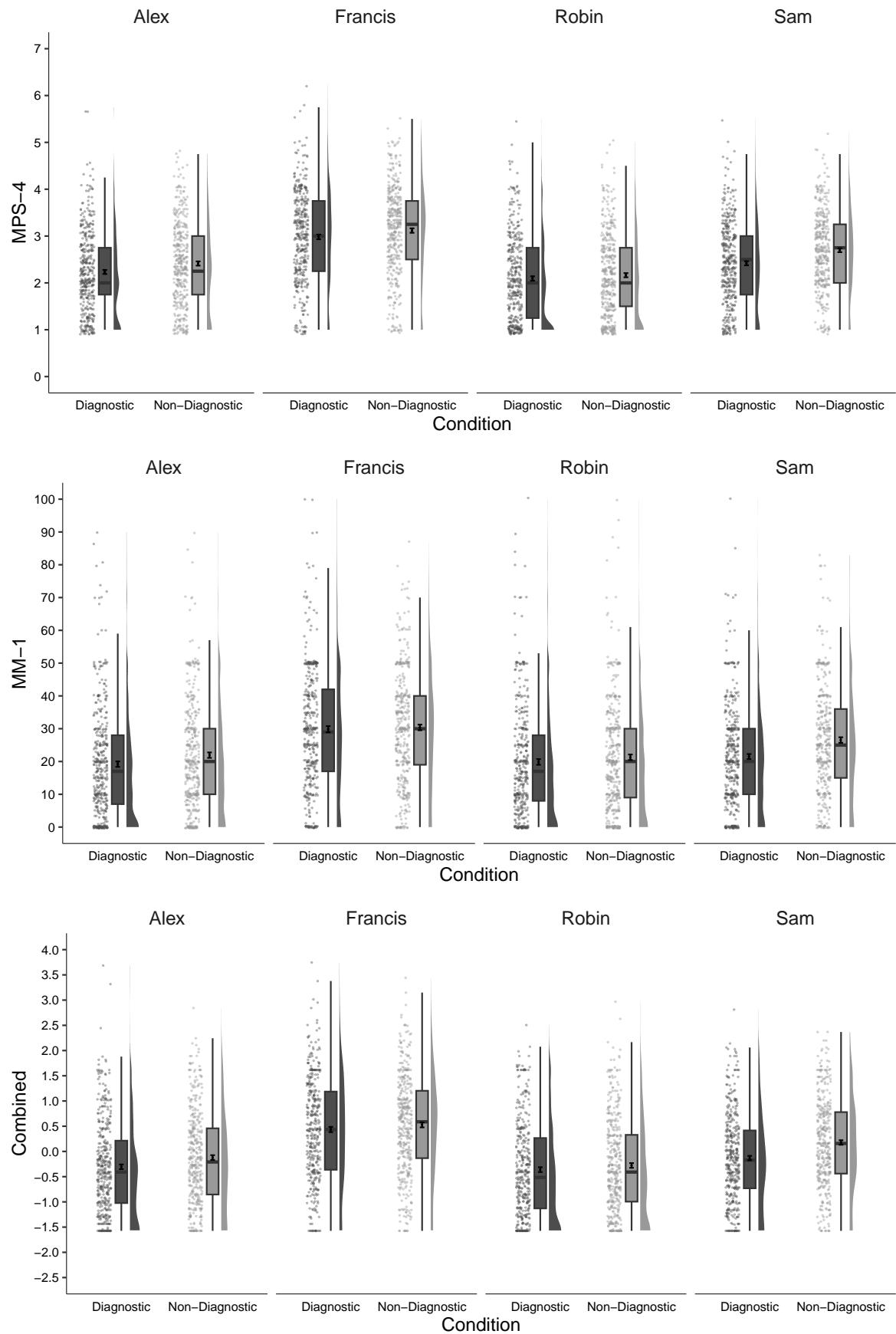


Figure 3. Study 1: Differences in moral perception for each description

129 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
130 condition ($M = 2.16, SD = 0.90$), than in the diagnostic condition ($M = 2.09, SD = 0.92$),
131 $t(793.94) = -1.09, p = .275, d = 0.08$; MM-1 ratings were similar in the non-diagnostic
132 condition ($M = 21.29, SD = 16.94$), and in the diagnostic condition ($M = 19.87, SD =$
133 17.17), $t(794.97) = -1.18, p = .239, d = 0.08$. For the combined measure ratings were also
134 similar in the non-diagnostic condition ($M = -0.28, SD = 0.94$), and in the diagnostic
135 condition ($M = -0.36, SD = 0.94$), $t(796.03) = -1.24, p = .217, d = 0.09$.

136 For *Alex*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
137 $= 2.41, SD = 0.88$), than in the diagnostic condition ($M = 2.23, SD = 0.86$), $t(796.97) =$
138 $-2.92, p = .004, d = 0.21$; MM-1 ratings were higher in the non-diagnostic condition ($M =$
139 $21.93, SD = 16.47$), than in the diagnostic condition ($M = 19.20, SD = 16.73$), $t(798.89) =$
140 $-2.33, p = .020, d = 0.16$. For the combined measure ratings were also higher in the
141 non-diagnostic condition ($M = -0.12, SD = 0.92$), than in the diagnostic condition ($M =$
142 $-0.30, SD = 0.92$), $t(798.40) = -2.82, p = .005, d = 0.20$.

143 For *Francis*, MPS-4 scores were significantly higher for the non-diagnostic condition
144 ($M = 3.12, SD = 0.95$), than in the diagnostic condition ($M = 2.98, SD = 0.97$), $t(796.12)$
145 $= -1.99, p = .047, d = 0.14$; MM-1 ratings were not significantly different in the
146 non-diagnostic condition ($M = 30.38, SD = 17.17$), than in the diagnostic condition ($M =$
147 $29.84, SD = 18.56$), $t(788.61) = -0.43, p = .668, d = 0.03$. For the combined measure
148 ratings were also similar in the non-diagnostic condition ($M = 0.53, SD = 0.98$), and in the
149 diagnostic condition ($M = 0.44, SD = 1.02$), $t(794.36) = -1.29, p = .198, d = 0.09$.

150

Study 2 (good): Supplementary Analyses

151 **Study 2: Combined Measure**

152 Below we report the results for the combined measure of moral perception. We
153 additionally report the effect of condition on responses to each description individually

154 The means and standard deviations for the combined measure for each scenario are
155 as follows: *Sam*, $M = 0.07$, $SD = 0.97$, *Francis*, $M = -0.17$, $SD = 1.06$, *Alex*, $M = 0.09$, SD
156 = 1.02, *Robin*, $M = 0.07$, $SD = 0.96$. There was significant variation depending on the
157 description, $F(3,2335) = 48.01$, $p < .001$, partial $\eta^2 = 0.01$. *Francis* appeared to be rated
158 as the less favorable than all other characters (all $ps < .001$), there were no differences
159 between *Sam*, *Robin*, and *Alex* (all $ps > .05$).

160 We conducted a linear-mixed-effects model to test if condition influenced moral
161 perception. Our outcome measure was the combined moral perception measure, our
162 predictor variable was condition; we allowed intercepts and the effect of condition to vary
163 across participants, and scenario was also included in the model. Overall, the model
164 significantly predicted participants responses, and provided a better fit for the data than
165 the baseline model, $\chi^2(8) = 142.42$, $p < .001$. Condition did not influence moral
166 perception, $F(1, 2,452.92) = 0.88$, $p = .349$; and was not a significant predictor in the
167 model when controlling for scenario, $b = -0.01$, $t(2,613.53) = -0.42$, $p = .673$, see Figure 4.

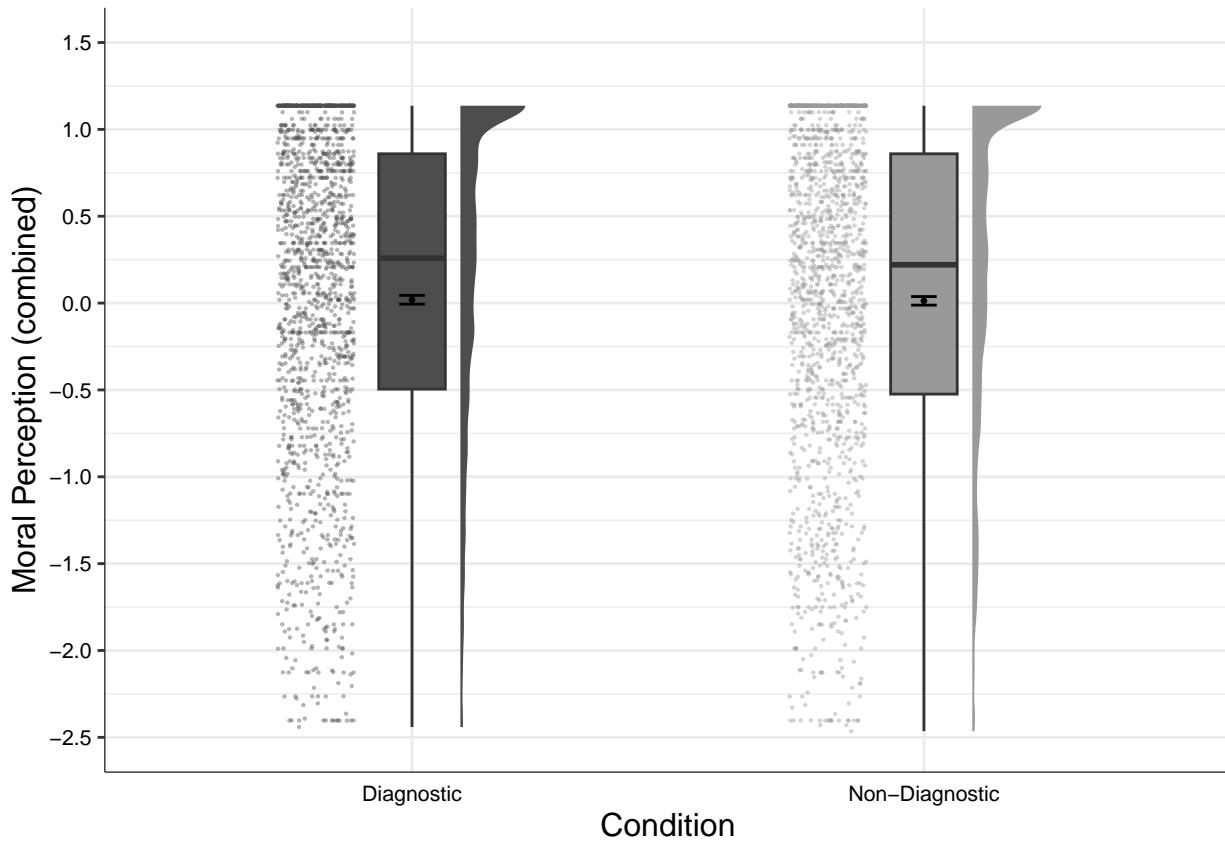


Figure 4. Study 2: Differences in combined measure depending on condition

¹⁶⁸ **Study 2: Differences between the Descriptions**

¹⁶⁹ Below we provide analyses of the effect of condition on responses to each scenario
¹⁷⁰ individually. The responses for each scenario across each measure depending on condition
¹⁷¹ are displayed in Figure 5.

¹⁷² For *Sam*, MPS-4 scores were not significantly different in the non-diagnostic condition
¹⁷³ ($M = 6.17$, $SD = 0.89$), than in the diagnostic condition ($M = 6.05$, $SD = 1.06$), $t(680.49)$
¹⁷⁴ = -1.71 , $p = .088$, $d = 0.12$; MM-1 ratings were similar in the non-diagnostic condition (M
¹⁷⁵ = 84.90 , $SD = 14.26$), and in the diagnostic condition ($M = 84.20$, $SD = 14.76$), $t(744.17)$
¹⁷⁶ = -0.69 , $p = .490$, $d = 0.05$. For the combined measure ratings were also similar in the
¹⁷⁷ non-diagnostic condition ($M = 0.11$, $SD = 0.93$), and in the diagnostic condition ($M =$
¹⁷⁸ 0.02 , $SD = 1.03$), $t(717.94) = -1.33$, $p = .183$, $d = 0.10$.

179 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
180 condition ($M = 6.08, SD = 1.00$), than in the diagnostic condition ($M = 6.13, SD = 0.98$),
181 $t(784.04) = 0.73, p = .463, d = 0.05$; MM-1 ratings were similar in the non-diagnostic
182 condition ($M = 84.12, SD = 14.37$), and in the diagnostic condition ($M = 85.98, SD =$
183 13.32), $t(800.09) = 1.92, p = .055, d = 0.13$. For the combined measure ratings were also
184 similar in the non-diagnostic condition ($M = 0.03, SD = 0.98$), and in the diagnostic
185 condition ($M = 0.13, SD = 0.95$), $t(788.76) = 1.46, p = .145, d = 0.10$.

186 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
187 condition ($M = 6.11, SD = 1.00$), than in the diagnostic condition ($M = 6.14, SD = 0.99$),
188 $t(737.60) = 0.32, p = .746, d = 0.02$; MM-1 ratings were similar in the non-diagnostic
189 condition ($M = 85.28, SD = 14.31$), and in the diagnostic condition ($M = 84.83, SD =$
190 15.51), $t(776.47) = -0.43, p = .668, d = 0.03$. For the combined measure ratings were also
191 similar in the non-diagnostic condition ($M = 0.09, SD = 0.98$), and in the diagnostic
192 condition ($M = 0.09, SD = 1.04$), $t(767.89) = -0.06, p = .952, d = 0.00$.

193 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
194 condition ($M = 5.82, SD = 1.05$), than in the diagnostic condition ($M = 5.90, SD = 1.08$),
195 $t(794.94) = 1.06, p = .290, d = 0.07$; MM-1 ratings were not significantly different in the
196 non-diagnostic condition ($M = 81.74, SD = 15.67$), than in the diagnostic condition ($M =$
197 $82.31, SD = 14.90$), $t(771.23) = 0.54, p = .591, d = 0.04$. For the combined measure
198 ratings were also similar in the non-diagnostic condition ($M = -0.20, SD = 1.08$), and in
199 the diagnostic condition ($M = -0.14, SD = 1.04$), $t(777.51) = 0.88, p = .379, d = 0.06$.

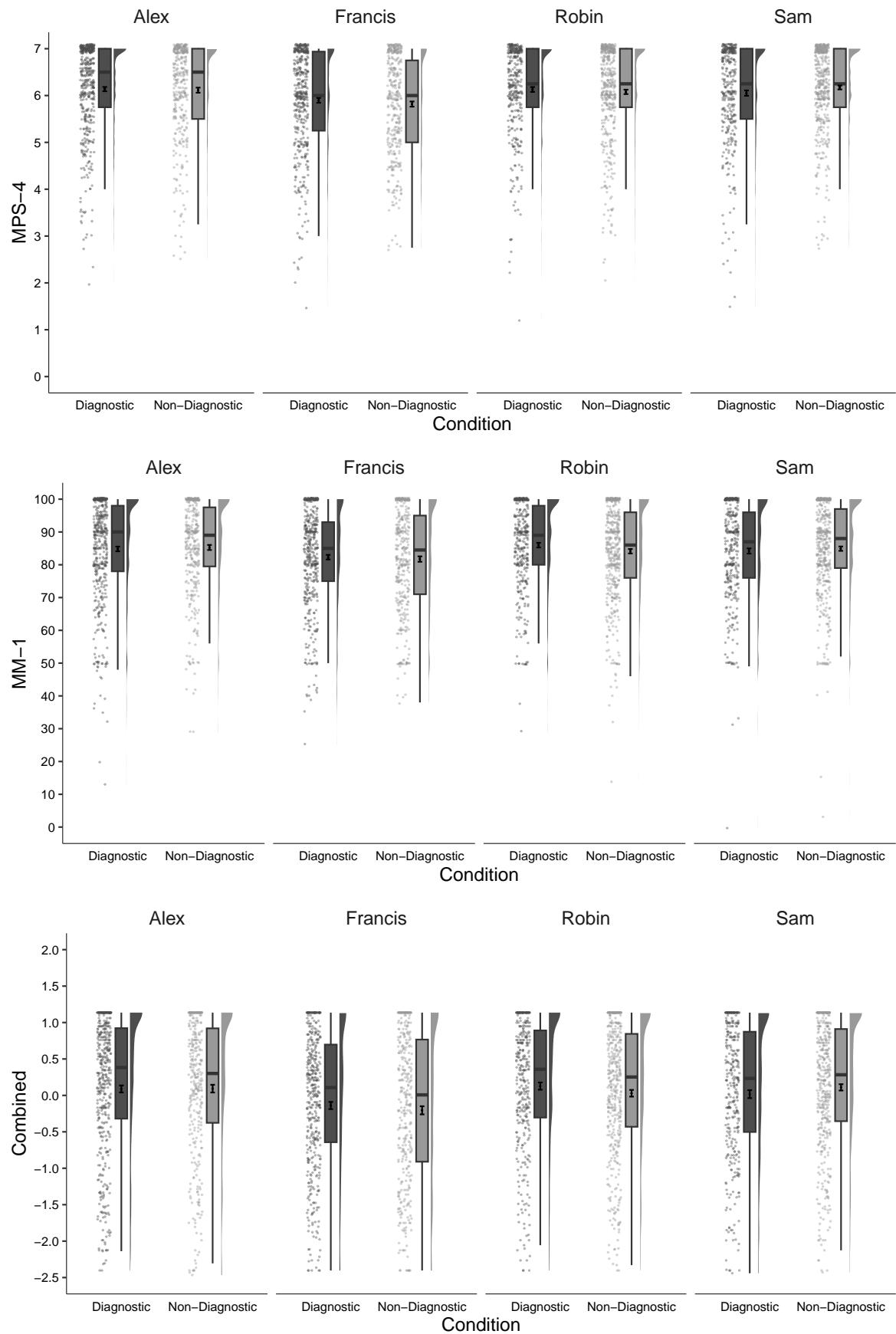


Figure 5. Study 2: Differences in moral perception for each description

200 Study 3 (bad and good): Supplementary Analyses**201 Study 3: Combined Measure**

202 Below we report the results for the combined measure of moral perception from both
203 DVs. We additionally report the effect of condition on responses to each description
204 individually

205 The means and standard deviations for the combined measure for each scenario are
206 as follows: *Sam*, $M = 0.93$, $SD = 0.39$, *Francis*, $M = -1.17$, $SD = 0.42$, *Alex*, $M = -1.08$,
207 $SD = 0.46$, *Robin*, $M = 0.99$, $SD = 0.36$. There was significant variation depending on the
208 description, $F(2,1403) = 6,772.79$, $p < .001$, partial $\eta^2 = 0.87$. Both the *good* characters
209 (*Robin* and *Sam*) were rated significantly more favorably than both the *bad* characters
210 (*Alex* and *Francis*; all $p < .001$). For the *good* characters, *Robin* was rated higher than
211 *Sam* ($p < .001$), and for the *bad* characters *Francis* was rated more negatively than *Alex* (p
212 $< .001$).

213 We conducted a linear-mixed-effects model to test if our predictors influenced
214 responses on the combined moral perception measure. Our outcome measure was the
215 combined moral perception measure, our predictor variables were condition and valence; we
216 allowed intercepts and the effects of condition and valence to vary across participants.
217 Overall, the model significantly predicted participants responses, and provided a better fit
218 for the data than the baseline model, $\chi^2(5) = 1,796.22$, $p < .001$. Condition significantly
219 influenced responses to the combined moral perception measure, $F(1, 828) = 47.25$, $p <$
220 $.001$ and was a significant predictor in the model when controlling for scenario, $b = -0.07$,
221 $t(827.54) = -6.87$, $p < .001$; valence significantly predicted responses, $F(1, 826) = 1,476.93$,
222 $p < .001$; and there was also a significant condition \times valence interaction, $F(1, 821) =$
223 4.23 , $p = .040$, see Figure 6.

224 For the *bad* characters, we conducted a linear-mixed-effects model to test if condition

225 influenced responses to the combined measure. Our outcome measure was the combined
226 moral perception measure, our predictor variable was condition; we allowed intercepts and
227 the effect of condition to vary across participants. Overall, the model significantly
228 predicted participants responses, and provided a better fit for the data than the baseline
229 model, $\chi^2(3) = 74.54, p < .001$. Condition significantly influenced MPS-4 responses $F(1,$
230 $820.39) = 37.63, p < .001$, and was a significant predictor in the model $b = -0.04, t(820.39)$
231 $= -6.13, p < .001$.

232 For the *good* characters, we conducted a linear-mixed-effects model to test if
233 condition influenced responses to the combined measure. Our outcome measure was the
234 combined moral perception measure, our predictor variable was condition; we allowed
235 intercepts and the effect of condition to vary across participants. Overall, the model
236 significantly predicted participants responses, and provided a better fit for the data than
237 the baseline model, $\chi^2(3) = 45.20, p < .001$. Condition significantly influenced MPS-4
238 responses $F(1, 826.21) = 15.67, p < .001$, and was a significant predictor in the model $b =$
239 $0.02, t(826.21) = 3.96, p < .001$.

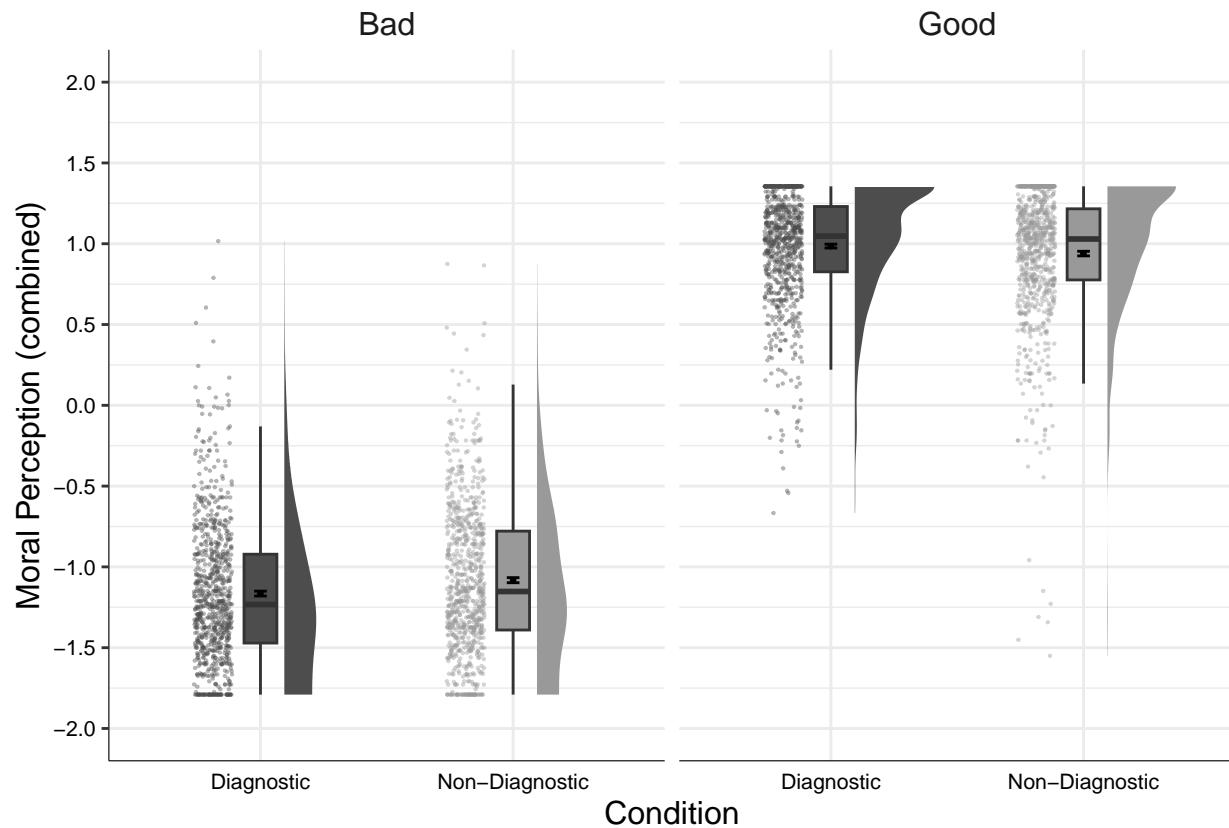


Figure 6. Study 3: Differences in the combined measure depending on condition

²⁴⁰ **Study 3: Differences between the descriptions**

²⁴¹ Again, we conducted separate analyses to investigate of condition on responses to
²⁴² each scenario individually. The responses for each scenario across each measure depending
²⁴³ on condition are displayed in Figure 7.

²⁴⁴ For *Sam*, MPS-4 scores were not significantly lower in the non-diagnostic condition
²⁴⁵ ($M = 6.15$, $SD = 0.86$), than in the diagnostic condition ($M = 6.25$, $SD = 0.76$), $t(812.83)$,
²⁴⁶ $= 1.68$, $p = .094$, $d = 0.12$; Similarly, MM-1 ratings were similar in the non-diagnostic
²⁴⁷ condition ($M = 85.49$, $SD = 14.10$), in the diagnostic condition ($M = 87.18$, $SD = 13.21$),
²⁴⁸ $t(821.76) = 1.78$, $p = .075$, $d = 0.12$. For the combined measure ratings was also no
²⁴⁹ significant difference between the non-diagnostic condition ($M = 0.90$, $SD = 0.42$), and the
²⁵⁰ diagnostic condition ($M = 0.96$, $SD = 0.37$), $t(811.12) = 1.88$, $p = .060$, $d = 0.13$.

251 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
252 condition ($M = 6.28, SD = 0.80$), than in the diagnostic condition ($M = 6.36, SD = 0.71$),
253 $t(809.44) = 1.60, p = .111, d = 0.11$; MM-1 ratings were similar in the non-diagnostic
254 condition ($M = 87.84, SD = 13.49$), and in the diagnostic condition ($M = 89.02, SD =$
255 10.30), $t(765.30) = 1.42, p = .156, d = 0.10$. For the combined measure ratings were also
256 similar in the non-diagnostic condition ($M = 0.97, SD = 0.39$), than in the diagnostic
257 condition ($M = 1.01, SD = 0.32$), $t(784.03) = 1.63, p = .103, d = 0.11$.

258 For *Alex*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
259 $= 2.41, SD = 0.88$), than in the diagnostic condition ($M = 2.24, SD = 0.90$), $t(830.38) =$
260 $-2.69, p = .007, d = 0.19$; MM-1 ratings were similar in the non-diagnostic condition ($M =$
261 $23.53, SD = 16.61$), and in the diagnostic condition ($M = 22.62, SD = 18.34$), $t(828.19) =$
262 $-0.75, p = .454, d = 0.05$. For the combined measure ratings were also similar in the
263 non-diagnostic condition ($M = -1.05, SD = 0.45$), and in the diagnostic condition ($M =$
264 $-1.11, SD = 0.47$), $t(830.90) = -1.77, p = .077, d = 0.12$.

265 For *Francis*, MPS-4 scores were significantly higher for the non-diagnostic condition
266 ($M = 2.26, SD = 0.85$), than in the diagnostic condition ($M = 2.05, SD = 0.70$), $t(802.80)$
267 $= -3.96, p < .001, d = 0.27$; MM-1 ratings were significantly higher in the non-diagnostic
268 condition ($M = 22.01, SD = 17.84$), than in the diagnostic condition ($M = 18.45, SD =$
269 15.76), $t(817.94) = -3.05, p = .002, d = 0.21$. For the combined measure ratings were also
270 significantly higher in the non-diagnostic condition ($M = -1.11, SD = 0.46$), than in the
271 diagnostic condition ($M = -1.23, SD = 0.38$), $t(808.55) = -3.85, p < .001, d = 0.27$.

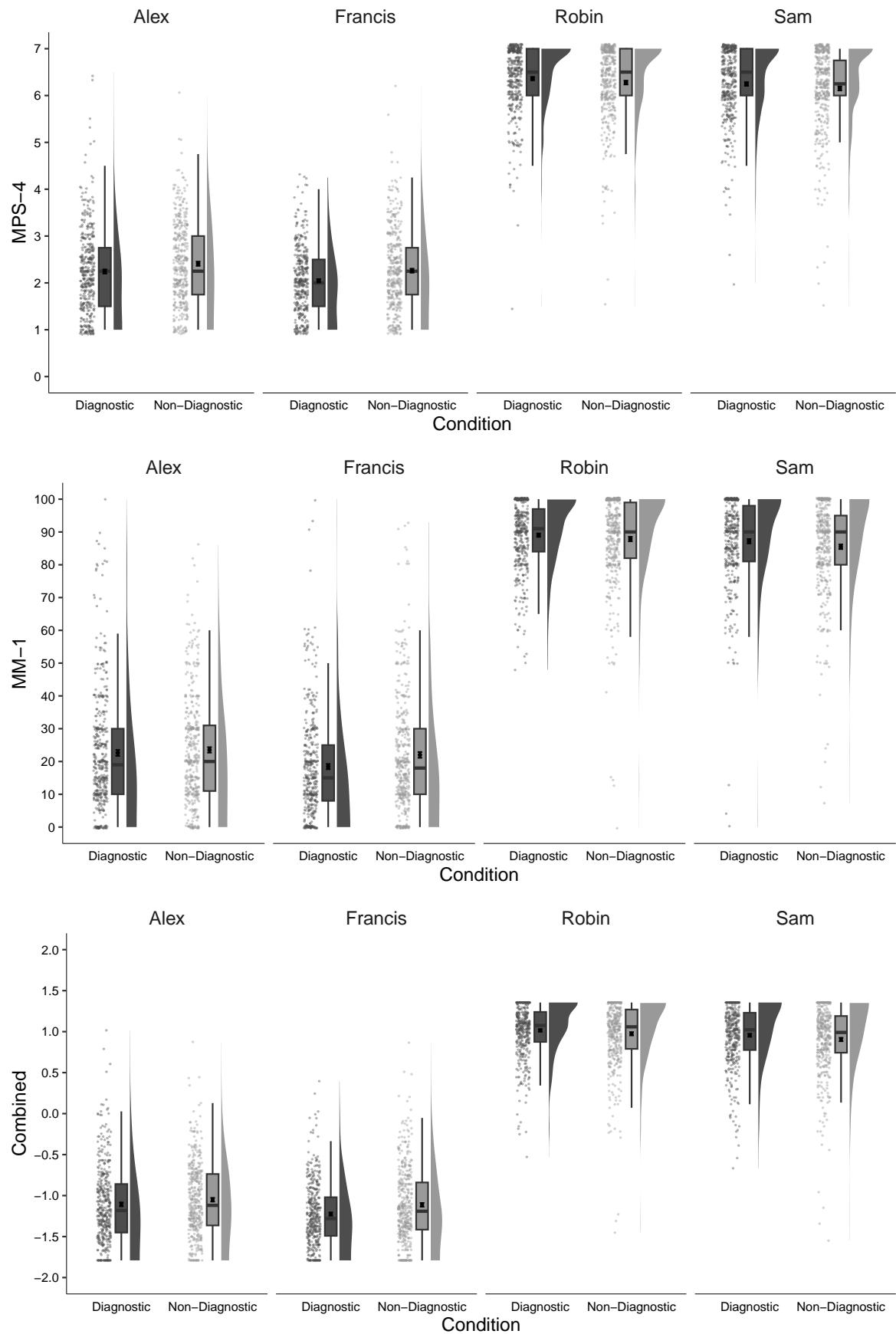


Figure 7. Study 3: Differences in moral perception for each description

272

Pilot Study 1

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

277 **Pilot Study 1: Method**

278 **Pilot 1: Participants and design.** The pilot study was a within-subjects design.

279 The independent variable was description type with two levels, *diagnostic* and
280 *non-diagnostic*. We used two dependent variables. The first dependent variable was the
281 four item moral perception scale (MPS-4), participants rated the characters on four
282 dimensions using 7-point bipolar scales. The dimensions and scale endpoints were:
283 Bad-Good, Immoral-Moral, Violent-Peaceful, Merciless-Empathetic, this showed excellent
284 reliability, $\alpha = 0.93$. The second dependent variable was a single item moral perception
285 measure (MM-1) which consisted of a 100-point slider ranging from 0 = *Very Immoral* to
286 100 = *Very Moral*. Both dependent variables were taken from Walker et al. (2021).

287 A total sample of 235 (89 female, 142 male, 1 non-binary, 1 prefer not to say; $M_{age} =$
288 36.45, min = 20, max = 72, $SD = 10.23$) started the survey. Participants were recruited
289 from MTurk.

290 We removed participants who failed both manipulation checks ($n = 23$), leaving a
291 total sample of 212 participants (80 female, 128 male, 1 non-binary, 1 prefer not to say;
292 $M_{age} = 36.63$, min = 20, max = 72, $SD = 10.34$).

293 **Pilot 1: Procedure and materials.** Data were collected using an online
294 questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented
295 with descriptions of six characters.

296 Moral character descriptions were developed by combining descriptions relating to

297 three different moral foundations. These descriptions were adapted from the items of the

298 extended character morality questionnaire (Grizzard et al., 2020), and read as follows:

299 (i) *Imagine a person named Sam. Throughout their life they have been known to be*
300 *cruel, act unfairly, and to betray their own group;*

301 (ii) *Imagine a person named Robin. Throughout their life they have been known to*
302 *physically hurt others, treat some people differently to others, and show lack of loyalty;*

303 (iii) *Imagine a person named Francis. Throughout their life they have been known to*
304 *violate the standards of purity and decency, show lack of respect for authority, and*
305 *treat people unequally*

306 (iv) *Imagine a person named Alex. Throughout their life they have been known to cause*
307 *others to suffer emotionally, to deny others their rights, and to cause chaos or*
308 *disorder.*

309 We developed neutral descriptions that included information relating to physical

310 appearance/attributes, hobbies/activities, and family information that read as follows:

311 (i) *Imagine a person named Jackie. They have red hair, play tennis four times a month,*
312 *and have one older sibling and one younger sibling;*

313 (ii) *Imagine a person named Charlie. They are left-handed, drink tea in the morning, and*
314 *have two older siblings and one younger sibling.*

315 Character descriptions did not specify the gender of the characters, and all characters

316 had names that could be either male or female (Sam, Robin, Francis, Alex, Jackie,

317 Charlie). All participants read six descriptions, four moral descriptions and two neutral.

318 Pilot Study 1 was pre-registered at https://aspredicted.org/3VK_8FD.

³¹⁹ **Pilot 1: Results**

³²⁰ **Pilot 1: Main Measures.** The means and standard deviations for MPS-4 for each
³²¹ scenario are as follows: *Sam* (diagnostic), $M_{MPS-4} = 4.35$, $SD_{MPS-4} = 1.90$, *Francis*
³²² (diagnostic), $M_{MPS-4} = 4.46$, $SD_{MPS-4} = 1.73$, *Alex* (diagnostic), $M_{MPS-4} = 4.44$, $SD_{MPS-4} =$
³²³ 1.79, *Robin* (diagnostic), $M_{MPS-4} = 4.35$, $SD_{MPS-4} = 1.96$, *Jackie* (non-diagnostic), M_{MPS-4}
³²⁴ = 5.40, $SD_{MPS-4} = 1.01$, *Charlie* (non-diagnostic), $M_{MPS-4} = 5.38$, $SD_{MPS-4} = 1.01$. For the
³²⁵ diagnostic descriptions, there was no significant variation depending on the description,
³²⁶ $F(3,600) = 1.58$, $p = .194$, partial $\eta^2 = 0.00$. For the non-diagnostic descriptions there was
³²⁷ no significant difference in ratings depending on description, $t(211) = -0.67$, $p = .506$, $d =$
³²⁸ 0.05.

³²⁹ The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
³³⁰ (diagnostic), $M_{MM-1} = 55.67$, $SD_{MM-1} = 30.47$; *Francis* (diagnostic), $M_{MM-1} = 58.22$,
³³¹ $SD_{MM-1} = 28.61$; *Alex* (diagnostic), $M_{MM-1} = 56.80$, $SD_{MM-1} = 29.45$; *Robin* (diagnostic),
³³² $M_{MM-1} = 55.49$, $SD_{MM-1} = 31.38$; *Jackie* (non-diagnostic), $M_{MM-1} = 73.00$, $SD_{MM-1} =$
³³³ 14.72; *Charlie* (non-diagnostic), $M_{MM-1} = 72.94$, $SD_{MM-1} = 14.79$. For the diagnostic
³³⁴ descriptions, we observed significant variation depending on the description, $F(3,608) =$
³³⁵ 3.01, $p = .032$, partial $\eta^2 = 0.001$. When correcting for multiple comparisons, pairwise
³³⁶ comparisons did not reveal significant differences between descriptions. We note that
³³⁷ without correction, *Francis* appeared to be rated as more moral than both *Robin* ($p =$
³³⁸ .012), and *Sam* ($p = .009$). For the non-diagnostic descriptions there was no significant
³³⁹ difference in ratings depending on description, $t(211) = -0.09$, $p = .929$, $d = 0.01$.

³⁴⁰ We conducted a linear-mixed-effects model to test if condition influenced MPS-4
³⁴¹ responses. Our outcome measure was MPS-4, our predictor variable was condition; we
³⁴² allowed intercepts and the effect of condition to vary across participants. Overall, the
³⁴³ model significantly predicted participants responses, and provided a better fit for the data
³⁴⁴ than the baseline model, $\chi^2(2) = 860.16$, $p < .001$. Condition was a significant predictor in

345 the model $b = -0.49$, $t(211.05) = -8.54$, $p < .001$, with the non-diagnostic descriptions
 346 being rated as more moral than the diagnostic descriptions of immoral characters Figure 8.

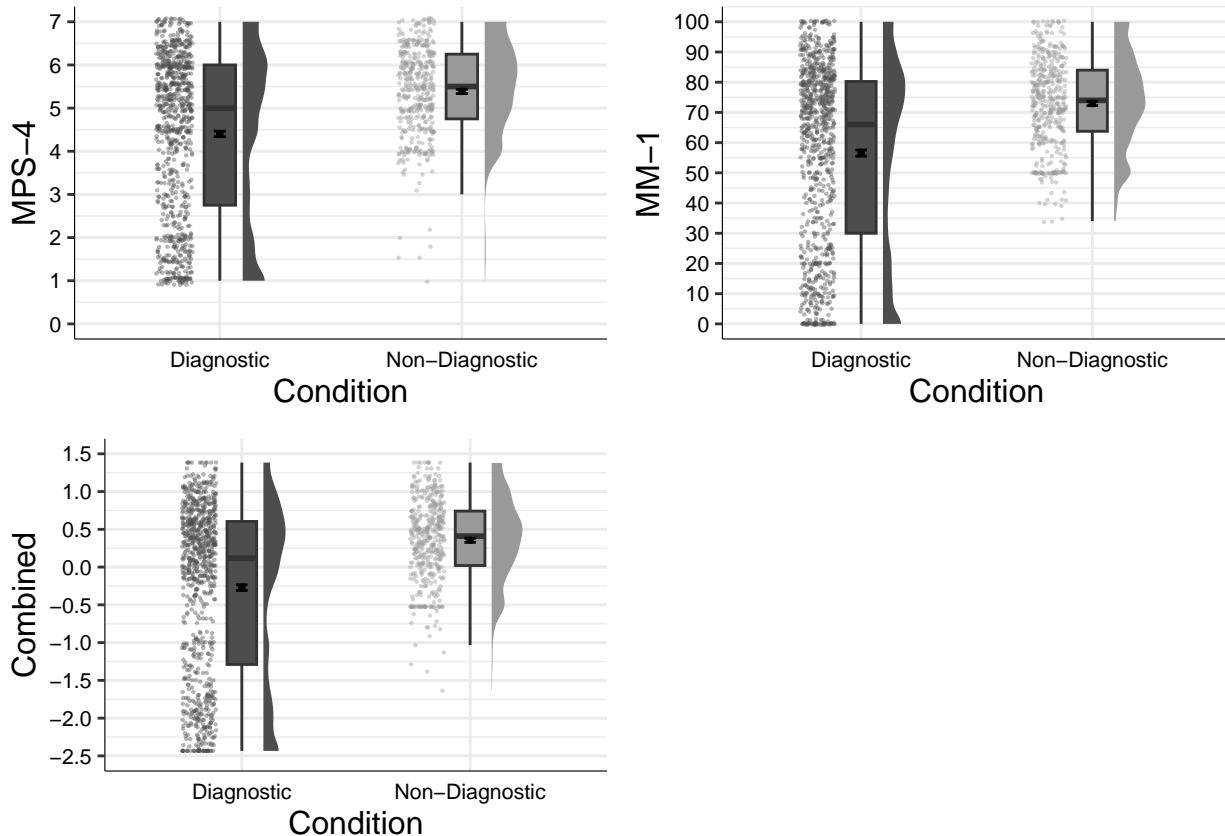


Figure 8. Pilot Study 1: Differences in moral perception depending on condition

347 We conducted a linear-mixed-effects model to test if condition influenced MM-1
 348 responses. Our outcome measure was MM-1, our predictor variable was condition; we
 349 allowed intercepts and the effect of condition to vary across participants. Overall, the
 350 model significantly predicted participants responses, and provided a better fit for the data
 351 than the baseline model, $\chi^2(2) = 924.82$, $p < .001$. Condition was a significant predictor in
 352 the model $b = -8.22$, $t(210.98) = -8.60$, $p < .001$, with the non-diagnostic descriptions
 353 being rated as more moral than the diagnostic descriptions, see Figure 8.

354 **Pilot 1: Combined Measure.** We developed a combined moral perception
 355 measure by calculating the mean of the combined mean-centered scores for MPS-4 and
 356 MM-1, and mean-centering this result. Below we report the analyses for this combined

357 measure.

358 The standardized means and standard deviations for the combined measure for each
359 scenario are as follows: *Sam* (diagnostic), $M = -0.30$, $SD = 1.16$; *Francis* (diagnostic), M
360 $= -0.22$, $SD = 1.06$; *Alex* (diagnostic), $M = -0.25$, $SD = 1.10$; *Robin* (diagnostic), $M =$
361 -0.31 , $SD = 1.19$; *Jackie* (non-diagnostic), $M = 0.36$, $SD = 0.55$; *Charlie* (non-diagnostic),
362 $M = 0.35$, $SD = 0.55$. For the moral descriptions, we observed significant variation
363 depending on the description, $F(3,602) = 2.67$, $p = .050$, partial $\eta^2 = 0.001$. When
364 correcting for multiple comparisons, pairwise comparisons did not reveal significant
365 differences between descriptions. We note that without correction, *Francis* appeared to be
366 rated as more moral than both *Robin* ($p = .022$), and *Sam* ($p = .021$). For the neutral
367 descriptions there was no significant difference in ratings depending on description, $t(211)$
368 $= -0.46$, $p = .645$, $d = 0.03$.

369 We conducted a linear-mixed-effects model to test if condition influenced responses
370 on this combined measure. Overall, the model significantly predicted participants
371 responses, and provided a better fit for the data than the baseline model $\chi^2(2) = 1,035.36$,
372 $p < .001$, and condition was a significant predictor in the model $b = -0.31$, $t(210.99) =$
373 -8.74 , $p < .001$. Participants rated the neutral/non-diagnostic descriptions as more moral
374 than the immoral/diagnostic descriptions (see Figure 8).

375

Pilot Study 2

376 Pilot Study 1 developed materials for studying the dilution effect with morally *bad*
377 characters. In Pilot Study 2, we develop materials for studying the dilution effect with
378 morally *good* characters. As with Pilot Study 1, we developed diagnostic and
379 non-diagnostic descriptions. We hypothesized that evaluations of the diagnostic
380 descriptions would be more extreme (more moral) than for the non-diagnostic descriptions

381 **Pilot Study 2: Method**

382 **Pilot 2: Participants and design.** The pilot study was a within-subjects design.
383 The independent variable was description type with two levels, *diagnostic* and
384 *non-diagnostic*. We used the same two dependent variables as in previous studies, the four
385 item moral perception scale (MPS-4, $\alpha = 0.84$), and the single item moral perception
386 measure (MM-1).

387 A total sample of 245 (70 female, 175 male, 0 non-binary, 0 prefer not to say; $M_{age} =$
388 36.69, min = 18, max = 71, $SD = 9.57$) started the survey. Participants were recruited
389 from MTurk.

390 We removed participants who failed both manipulation checks ($n = 30$), leaving a
391 total sample of 215 participants (63 female, 152 male, 0 non-binary, 0 prefer not to say;
392 $M_{age} = 36.59$, min = 18, max = 71, $SD = 9.59$).

393 **Pilot 2: Procedure and materials.** Data were collected using an online
394 questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented
395 with descriptions of six characters.

396 Moral character descriptions were developed by combining descriptions relating to
397 three different moral foundations, focusing on upholding the moral foundations (rather
398 than transgressions as in previous studies). We developed 4 descriptions of moral
399 characters that read as follows:

- 400 (i) *Imagine a person named Sam. Throughout their life they have been known to always*
401 *help and care for others, treat everyone fairly and equally, and show a strong sense of*
402 *loyalty to others;*
- 403 (ii) *Imagine a person named Robin. Throughout their life they have been known to show*
404 *compassion and empathy for others, act with a sense of fairness and justice, and,*
405 *never to break their word;*
- 406 (iii) *Imagine a person named Francis. Throughout their life they have been known to*
407 *uphold the standards of purity and decency, show respect for authority, and to always*
408 *act honestly and fairly;*
- 409 (iv) *Imagine a person named Alex. Throughout their life they have been known to protect*
410 *and provide shelter to the weak and vulnerable, uphold the rights of others, and show*
411 *respect for authority.*

412 We developed 2 descriptions of morally neutral characters that included information

413 relating to physical appearance/attributes, hobbies/activities, and a color preference:

- 414 (i) *Imagine a person named Jackie. They have dark hair, go for a jog twice a week, and*
415 *their favourite colour is blue;*
- 416 (ii) *Imagine a person named Charlie. They have blue eyes, drink coffee in the morning,*
417 *and their favourite colour is green.*

418 We used the same gender ambiguous names, and we did not specify the gender of the

419 characters. Pilot Study 2 was pre-registered at https://aspredicted.org/W52_VPX.

420 **Pilot 2: Results**

421 **Pilot 2: Main Measures.** The means and standard deviations for MPS-4 for each

422 scenario are as follows: *Sam* (diagnostic), $M_{MPS-4} = 6.01$, $SD_{MPS-4} = 0.91$, *Francis*
423 (diagnostic), $M_{MPS-4} = 5.89$, $SD_{MPS-4} = 0.95$, *Alex* (diagnostic), $M_{MPS-4} = 5.94$, $SD_{MPS-4} =$

424 0.94, *Robin* (diagnostic), $M_{\text{MPS-4}} = 5.93$, $SD_{\text{MPS-4}} = 0.92$, *Jackie* (non-diagnostic), $M_{\text{MPS-4}}$
 425 $= 5.60$, $SD_{\text{MPS-4}} = 0.99$, *Charlie* (non-diagnostic), $M_{\text{MPS-4}} = 5.53$, $SD_{\text{MPS-4}} = 1.08$. For the
 426 diagnostic descriptions, there was significant variation depending on the description,
 427 $F(3,613) = 2.91$, $p = .036$, partial $\eta^2 = 0.00$, *Sam* was viewed significantly more favorably
 428 than *Francis* ($p = .040$). For the non-diagnostic descriptions there was no significant
 429 difference in ratings depending on description, $t(214) = -1.79$, $p = .075$, $d = 0.12$.

430 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
 431 (*diagnostic*), $M_{\text{MM-1}} = 79.85$, $SD_{\text{MM-1}} = 15.44$; *Francis* (*diagnostic*), $M_{\text{MM-1}} = 78.30$,
 432 $SD_{\text{MM-1}} = 15.84$; *Alex* (*diagnostic*), $M_{\text{MM-1}} = 79.78$, $SD_{\text{MM-1}} = 15.71$; *Robin* (*diagnostic*),
 433 $M_{\text{MM-1}} = 79.46$, $SD_{\text{MM-1}} = 15.41$; *Jackie* (*non-diagnostic*), $M_{\text{MM-1}} = 73.44$, $SD_{\text{MM-1}} =$
 434 15.83; *Charlie* (*non-diagnostic*), $M_{\text{MM-1}} = 73.07$, $SD_{\text{MM-1}} = 16.22$. For the diagnostic
 435 descriptions, we observed no significant variation depending on the description, $F(3,594) =$
 436 1.45, $p = .231$, partial $\eta^2 = 0.002$. For the non-diagnostic descriptions there was no
 437 significant difference in ratings depending on description, $t(214) = -0.60$, $p = .552$, $d =$
 438 0.04.

439 We conducted a linear-mixed-effects model to test if condition influenced MPS-4
 440 responses. Our outcome measure was MPS-4, our predictor variable was condition; we
 441 allowed intercepts and the effect of condition to vary across participants. Overall, the
 442 model significantly predicted participants responses, and provided a better fit for the data
 443 than the baseline model, $\chi^2(2) = 475.42$, $p < .001$. Condition was a significant predictor in
 444 the model $b = 0.19$, $t(214.35) = 6.53$, $p < .001$, with the diagnostic descriptions being
 445 rated as more moral than the non-diagnostic descriptions of immoral characters Figure 9.

446 We conducted a linear-mixed-effects model to test if condition influenced MM-1
 447 responses. Our outcome measure was MM-1, our predictor variable was condition; we
 448 allowed intercepts and the effect of condition to vary across participants. Overall, the
 449 model significantly predicted participants responses, and provided a better fit for the data

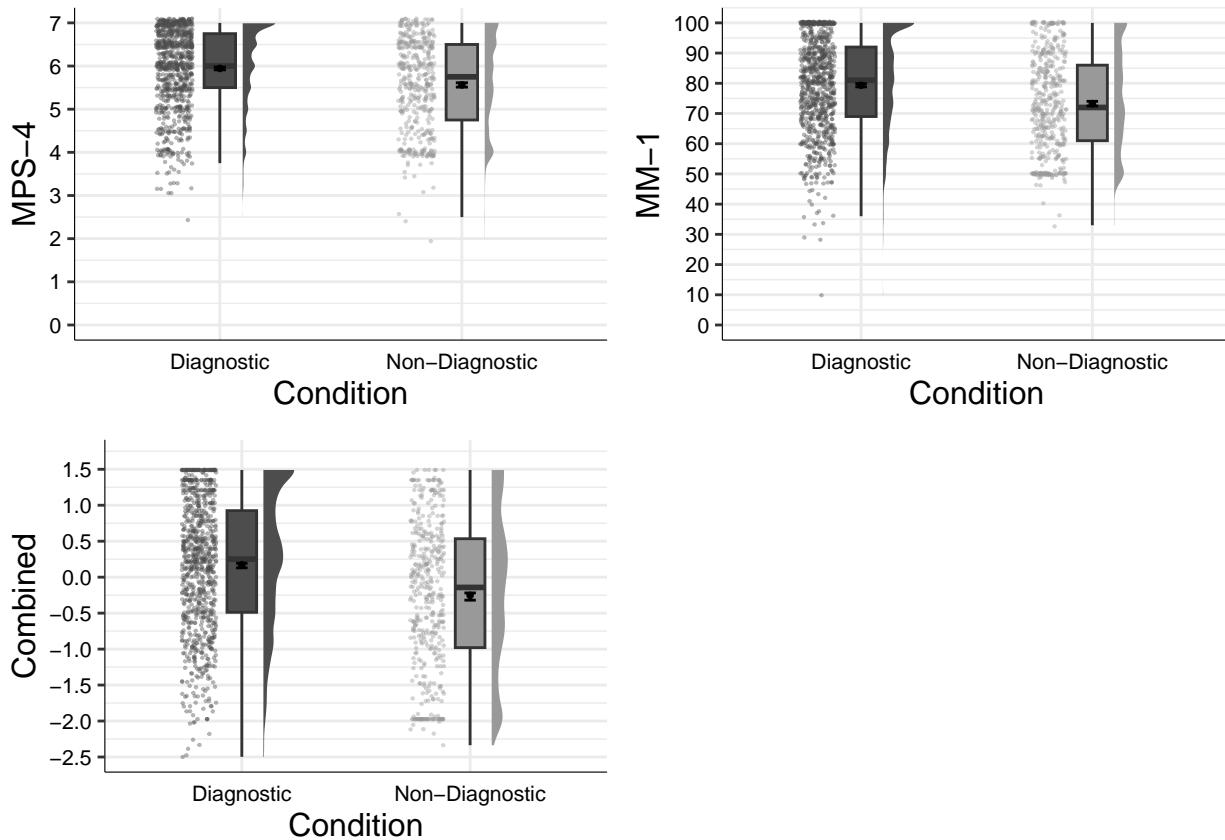


Figure 9. Pilot Study 2: Differences in moral perception depending on condition

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 9.

Pilot 2: Combined Measure. As in previous studies, we developed a combined moral perception measure by calculating the mean of the combined mean-centered scores for MPS-4 and MM-1, and mean-centering this result. Below we report the analyses for this combined measure.

The standardized means and standard deviations for the combined measure for each scenario are as follows: *Sam* (moral), $M = 0.21, SD = 0.91$; *Francis* (moral), $M = 0.10, SD = 0.96$; *Alex* (moral), $M = 0.18, SD = 0.94$; *Robin* (moral), $M = 0.16, SD = 0.93$; *Jackie* (neutral), $M = -0.24, SD = 1.01$; *Charlie* (neutral), $M = -0.30, SD = 1.07$. For the moral

descriptions, we observed significant variation depending on the description, $F(3,588) = 2.90$, $p = .039$, partial $\eta^2 = 0.002$. *Sam* was viewed significantly more favorably than *Francis* ($p = .045$). For the neutral descriptions there was no significant difference in ratings depending on description, $t(426.74) = -0.51$, $p = .609$, $d = 0.10$.

We conducted a linear-mixed-effects model to test if condition influenced responses to the combined measure. Overall, the model significantly predicted participants responses, and provided a better fit for the data than the baseline model $\chi^2(2) = 564.98$, $p < .001$, and condition was a significant predictor in the model $b = 0.22$, $t(214.32) = 6.60$, $p < .001$ (see Figure 9).

470 Study S1 - Good Characters

471 Study S1 is a replication of Study 2, but with an MTurk Sample.

472 **Study S1: Method**

473 **Study S1: Participants and design.** The design, materials, and procedure for
474 Study S1 were the same as for Study 2, the only change from Study 2 was that all
475 participants in Study S2 were recruited from MTurk. Study S1 was a within-subjects
476 design. The independent variable was condition with two levels, diagnostic and
477 non-diagnostic. We used the same two dependent variables as in previous studies, the four
478 item moral perception scale (MPS-4, $\alpha = 0.81$), and the single item moral perception
479 measure MM-1.

480 A total sample of 1118 (445 female, 642 male, 2 non-binary, 3 other; 1 prefer not to
481 say, $M_{age} = 37.44$, min = 19, max = 84, $SD = 11.08$) started the survey. Participants were
482 recruited from MTurk and paid \$0.40 for their participation.

483 Participants who failed both manipulation checks were removed ($n = 262$), leaving a
484 total sample of 856 participants (347 female, 507 male, 0 other, 0 prefer not to say; $M_{age} =$
485 37.12, min = 19, max = 84, $SD = 11.04$).

486 **Study S1: Procedure and materials.** All materials and procedures were the
487 same as in Study 2.

488 **Study S1: Results**

489 **Study S1: Main Measures.** The means and standard deviations for MPS-4 for
490 each scenario are as follows: *Sam*, $M_{MPS-4} = 5.95$, $SD_{MPS-4} = 0.93$, *Francis*, $M_{MPS-4} = 5.89$,
491 $SD_{MPS-4} = 0.91$, *Alex*, $M_{MPS-4} = 5.94$, $SD_{MPS-4} = 0.96$, *Robin*, $M_{MPS-4} = 5.95$, $SD_{MPS-4} =$
492 0.94. There was significant variation depending on the description, $F(3,2527) = 3.30$, $p =$

⁴⁹³ .020, partial $\eta^2 = 0.001$. Pairwise comparisons did not reveal any significant differences
⁴⁹⁴ between individual descriptions (all $ps > .05$).

⁴⁹⁵ The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
⁴⁹⁶ (diagnostic/moral), $M_{MM-1} = 81.34$, $SD_{MM-1} = 14.14$; *Francis* (diagnostic/moral), M_{MM-1}
⁴⁹⁷ = 80.65, $SD_{MM-1} = 14.16$; *Alex* (diagnostic/moral), $M_{MM-1} = 81.15$, $SD_{MM-1} = 14.42$;
⁴⁹⁸ *Robin* (diagnostic/moral), $M_{MM-1} = 81.63$, $SD_{MM-1} = 14.15$. There was significant
⁴⁹⁹ variation depending on the description, $F(3,2518) = 2.89$, $p = .035$, partial $\eta^2 = 0.001$.
⁵⁰⁰ Pairwise comparisons did not reveal any significant differences between individual
⁵⁰¹ descriptions (all $ps > .05$).

⁵⁰² We conducted a linear-mixed-effects model to test if condition influenced MPS-4
⁵⁰³ responses. Our outcome measure was MPS-4, our predictor variable was condition; we
⁵⁰⁴ allowed intercepts and the effect of condition to vary across participants, and scenario was
⁵⁰⁵ also included in the model. Overall, the model significantly predicted participants
⁵⁰⁶ responses, and provided a better fit for the data than the baseline model, $\chi^2(8) = 17.86$, p
⁵⁰⁷ = .022. Condition did not influence responses to the MPS-4, $F(1, 866.60) = 2.80$, $p = .095$;
⁵⁰⁸ and was not a significant predictor in the model when controlling for scenario, $b = 0.01$,
⁵⁰⁹ $t(867) = 1.67$, $p = .095$, see Figure 10.

⁵¹⁰ We conducted a linear-mixed-effects model to test if condition influenced MM-1
⁵¹¹ responses. Our outcome measure was MM-1, our predictor variable was condition; we
⁵¹² allowed intercepts and the effect of condition to vary across participants. Overall, the
⁵¹³ model significantly predicted participants responses, and provided a better fit for the data
⁵¹⁴ than the baseline model, $\chi^2(8) = 40.10$, $p < .001$. Condition significantly influenced MM-1
⁵¹⁵ responses $F(1, 864) = 4.79$, $p = .029$, and was a significant predictor in the model $b = 0.29$,
⁵¹⁶ $t(864) = 2.19$, $p = .029$, see Figure 10.

⁵¹⁷ **Study S1: Combined Measure.** The means and standard deviations for the
⁵¹⁸ combined measure for each scenario are as follows: *Sam*, $M = 0.03$, $SD = 1.02$, *Francis*, M

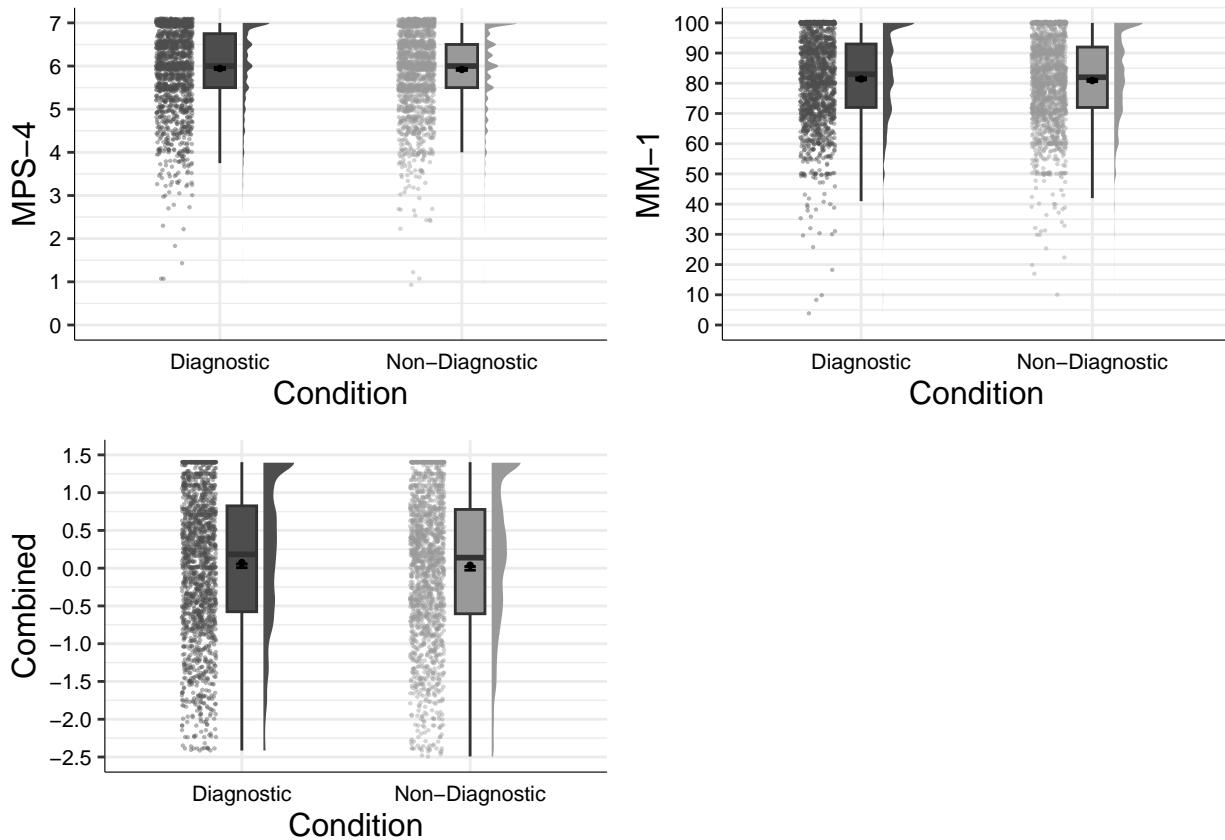


Figure 10. Study S1: Responses to moral perception measures depending on condition

519 $= -0.03$, $SD = 0.98$, *Alex*, $M = 0.02$, $SD = 1.04$, *Robin*, $M = 0.04$, $SD = 1.01$. There was
 520 significant variation depending on the description, $F(3,2493) = 4.32$, $p = .005$, partial $\eta^2 =$
 521 0.00. Follow-up pairwise comparisons did not reveal any significant differences between the
 522 different characters (all $ps > .05$).

523 We conducted a linear-mixed-effects model to test if condition influenced moral
 524 perception. Our outcome measure was the combined moral perception measure, our
 525 predictor variable was condition; we allowed intercepts and the effect of condition to vary
 526 across participants, and scenario was also included in the model. Overall, the model
 527 significantly predicted participants responses, and provided a better fit for the data than
 528 the baseline model, $\chi^2(8) = 42.42$, $p < .001$. Condition did not influence moral perception,
 529 $F(1, 865.01) = 5.31$, $p = .021$; and was not a significant predictor in the model when
 530 controlling for scenario, $b = -0.01$, $t(2,541.03) = -0.82$, $p = .410$, see Figure 4.

Study S1: Differences between the Descriptions. Below we provide analyses

of the effect of condition on responses to each scenario individually. The responses for each scenario across each measure depending on condition are displayed in Figure 11.

For *Sam*, MPS-4 scores were not significantly different in the non-diagnostic condition

($M = 5.89$, $SD = 0.91$), than in the diagnostic condition ($M = 6.02$, $SD = 0.95$), $t(810.53)$

$= 1.97$, $p = .049$, $d = 0.14$; MM-1 ratings were similar in the non-diagnostic condition (M

$= 79.75$, $SD = 14.62$), than in the diagnostic condition ($M = 83.25$, $SD = 13.30$), $t(845.88)$

$= 3.66$, $p < .001$, $d = 0.25$. For the combined measure ratings were also similar in the

non-diagnostic condition ($M = -0.06$, $SD = 1.03$), than in the diagnostic condition ($M =$

0.15 , $SD = 1.01$), $t(829.20) = 3.07$, $p = .002$, $d = 0.21$.

For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic

condition ($M = 5.95$, $SD = 0.93$), than in the diagnostic condition ($M = 5.94$, $SD = 0.95$),

$t(811.83) = -0.20$, $p = .841$, $d = 0.01$; MM-1 ratings were similar in the non-diagnostic

condition ($M = 81.62$, $SD = 14.28$), and in the diagnostic condition ($M = 81.64$, $SD =$

14.02), $t(824.54) = 0.02$, $p = .982$, $d = 0.00$. For the combined measure ratings were also

similar in the non-diagnostic condition ($M = 0.04$, $SD = 1.03$), than in the diagnostic

condition ($M = 0.04$, $SD = 0.99$), $t(828.47) = -0.10$, $p = .919$, $d = 0.01$.

For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic

condition ($M = 5.97$, $SD = 0.91$), than in the diagnostic condition ($M = 5.91$, $SD = 0.99$),

$t(845.29) = -0.91$, $p = .362$, $d = 0.06$; MM-1 ratings were similar in the non-diagnostic

condition ($M = 81.93$, $SD = 13.38$), than in the diagnostic condition ($M = 80.51$, $SD =$

15.21), $t(850.53) = -1.46$, $p = .145$, $d = 0.10$. For the combined measure ratings were also

similar in the non-diagnostic condition ($M = 0.07$, $SD = 0.98$), than in the diagnostic

condition ($M = -0.02$, $SD = 1.09$), $t(847.27) = -1.30$, $p = .192$, $d = 0.09$.

For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic

condition ($M = 5.87$, $SD = 0.95$), than in the diagnostic condition ($M = 5.91$, $SD = 0.87$),

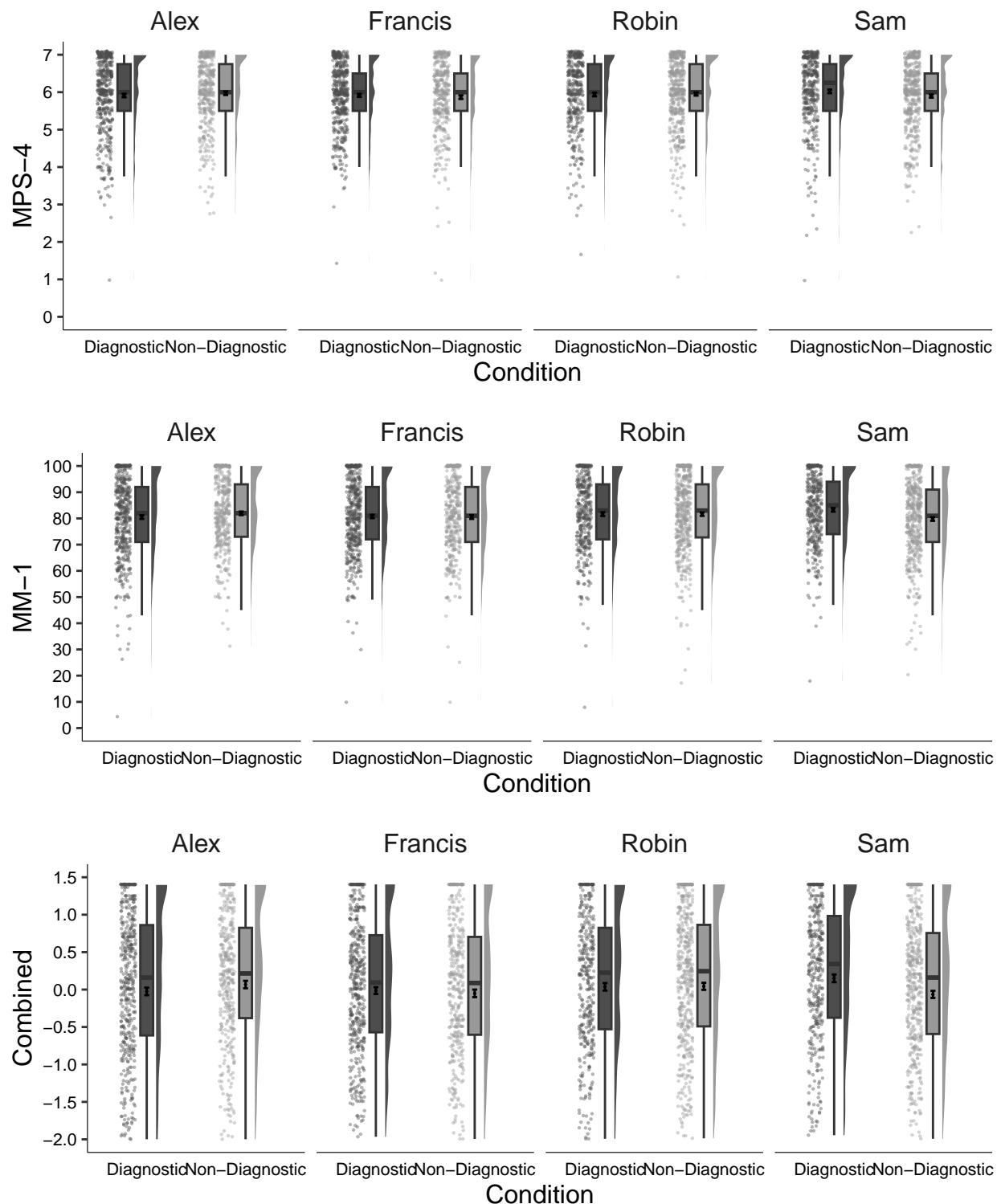


Figure 11. Study 2: Differences in moral perception for each description

557 $t(787.36) = 0.77, p = .443, d = 0.05$; MM-1 ratings were not significantly different in the
558 non-diagnostic condition ($M = 80.54, SD = 14.38$), than in the diagnostic condition ($M =$
559 $80.75, SD = 13.99$), $t(809.63) = 0.21, p = .832, d = 0.01$. For the combined measure
560 ratings were also similar in the non-diagnostic condition ($M = -0.05, SD = 0.99$), and in
561 the diagnostic condition ($M = -0.01, SD = 0.98$), $t(814.30) = 0.55, p = .581, d = 0.04$.

562

Study S2 - Good and Bad Characters

563

Study S2 is the same as Study 3, but with an MTurk sample. Study 3 was

564

pre-registered at https://aspredicted.org/QDF_XT1.

565

Study S2: Method

566

Study S2: Participants and design. Study S2 was a 2×2 within-subjects

567

factorial design. The first independent variable was condition with two levels, diagnostic

568

and non-diagnostic. The second independent variable was valence of character description,

569

with two levels morally good and morally bad. We used the same two dependent variables

570

as in previous studies, the four item moral perception scale (MPS-4, $\alpha = 0.94$), and the

571

single item moral perception measure MM-1.

572

A total sample of 1095 (386 female, 700 male, 2 non-binary, 0 other; 2 prefer not to

573

say, $M_{age} = 36.42$, min = 19, max = 77, $SD = 10.65$) started the survey. Participants were

574

recruited from MTurk and paid \$0.40 for their participation.

575

Participants who failed both manipulation checks were removed ($n = 221$), leaving a

576

total sample of 874 participants (320 female, 550 male, 0 other, 0 prefer not to say; $M_{age} =$

577

36.37, min = 19, max = 77, $SD = 10.72$).

578

Study S2: Procedure and materials. Again, data were collected using an online

579

questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented

580

with four descriptions of characters as in Study 3. To ensure consistency across character

581

judgments, we selected descriptions that related to the same moral foundations (care,

582

fairness, and loyalty). We used the same four character names as in previous studies. The

583

good characters were *Sam* and *Robin*, and the bad characters were *Francis* and *Alex*, e.g.,

584

Imagine a person named Robin. Throughout their life they have been known to show

585

compassion and empathy for others, act with a sense of fairness and justice, and, never to

586

break their word. or, *Imagine a person named Alex. Throughout their life they have been*

587 known to be cruel, act unfairly, and to betray their own group. Full descriptions for each
588 character are in the supplementary materials. One description for each the *good* and *bad*
589 characters was randomly assigned to include non-diagnostic information for each
590 participant thus all participants were exposed to all conditions (see
591 https://osf.io/mdnpv/?view_only=77883e3fb3d45f1a35fe92d5318cb67for details of the
592 randomization blocks). Study S2 was pre-registered at https://aspredicted.org/QDF_XT1

593 **Study S2: Results**

594 The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*
595 (*good*), $M_{MPS-4} = 5.90$, $SD_{MPS-4} = 1.03$, *Francis* (*bad*), $M_{MPS-4} = 4.07$, $SD_{MPS-4} = 2.07$,
596 *Alex* (*bad*), $M_{MPS-4} = 4.03$, $SD_{MPS-4} = 2.03$, *Robin* (*good*), $M_{MPS-4} = 5.85$, $SD_{MPS-4} =$
597 1.05. There was significant variation depending on the description, $F(1,1080) = 442.71$, p
598 $< .001$, partial $\eta^2 = 0.24$. Both the *good* characters (*Robin* and *Sam*) were rated
599 significantly more favorably than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$).
600 There were no differences between *Robin* and *Sam* (*good*: $p = .366$) or between *Alex* and
601 *Francis* (*bad*; $p = .648$).

602 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
603 (*good*), $M_{MM-1} = 81.01$, $SD_{MM-1} = 15.23$; *Francis* (*bad*), $M_{MM-1} = 51.49$, $SD_{MM-1} = 33.18$;
604 *Alex* (*bad*), $M_{MM-1} = 50.89$, $SD_{MM-1} = 32.14$; *Robin* (*good*), $M_{MM-1} = 80.81$, $SD_{MM-1} =$
605 15.16. There was significant variation depending on the description, $F(1,1080) = 458.92$, p
606 $< .001$, partial $\eta^2 = 0.254$. Again, the *good* characters (*Robin* and *Sam*) were rated
607 significantly more favorably than the *bad* characters (*Alex* and *Francis*; all $ps < .001$).
608 There were no differences between *Robin* and *Sam* (*good*: $p = .776$) or between *Alex* and
609 *Francis* (*bad*; $p = .683$).

610 We conducted a linear-mixed-effects model to test if our predictors influenced MPS-4
611 responses. Our outcome measure was MPS-4, our predictor variables were condition and

612 valence; we allowed intercepts and the effects of condition and valence to vary across
613 participants. Overall, the model significantly predicted participants responses, and
614 provided a better fit for the data than the baseline model, $\chi^2(5) = 4,554.31, p < .001$.
615 Overall, there was a significant main effect for condition, $F(1, 873) = 8.61, p = .003$;
616 valence significantly predicted responses, $F(1, 873) = 1,859.34, p < .001$; and there was no
617 significant condition \times valence interaction, $F(1, 873) = 0.01, p = .935$.

618 We conducted a linear-mixed-effects model to test if our predictors influenced MM-1
619 responses. The model was the same as the previous model, with a change to the outcome
620 measure, our outcome measure for this model was MM-1. As above, our predictor variables
621 were condition and valence; we allowed intercepts and the effects of condition and valence
622 to vary across participants. Overall, the model significantly predicted participants
623 responses, and provided a better fit for the data than the baseline model, $\chi^2(5) = 3,496.86$,
624 $p < .001$. Overall there was a main effect for condition, $F(1, 873) = 16.61, p < .001$;
625 valence significantly predicted responses, $F(1, 873) = 986.37, p < .001$; and there was no
626 significant condition \times valence interaction, $F(1, 873) = 0.04, p = .849$.

627 We conducted a linear-mixed-effects model to test if our predictors influenced
628 responses on the combined moral perception measure. Our outcome measure was the
629 combined moral perception measure, our predictor variables were condition and valence; we
630 allowed intercepts and the effects of condition and valence to vary across participants.
631 Overall, the model significantly predicted participants responses, and provided a better fit
632 for the data than the baseline model, $\chi^2(5) = 4,467.15, p < .001$. Condition significantly
633 influenced responses to the combined moral perception measure, $F(1, 873) = 16.65, p <$
634 $.001$ and was a significant predictor in the model when controlling for scenario, $b = -0.02$,
635 $t(873.00) = -4.08, p < .001$; valence significantly predicted responses, $F(1, 873) = 1,598.27$,
636 $p < .001$; and there was also a significant condition \times valence interaction, $F(1, 873) =$
637 $0.03, p = .867$, see Figure 6.

638 For both MP-4 and MM-1 (and the combined measure) we found a main effect for

639 condition and valence, and there was no condition \times valence interaction. We conducted

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

641 descriptions separately.

642 Differences in the *Bad* Descriptions

643 For the *bad* characters, we conducted a linear-mixed-effects model to test if condition

644 influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was

645 condition; we allowed intercepts and the effect of condition to vary across participants.

646 Overall, the model did not significantly predict participants responses, or provide a better

647 fit for the data than the baseline model, $\chi^2(3) = 5.40, p = .145$. Condition did not

648 significantly influence MPS-4 responses $F(1, 872.00) = 3.54, p = .060$, and was not a

649 significant predictor in the model $b = -0.03, t(872.00) = -1.88, p = .060$, see Figure 12.

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

651 responses. Our outcome measure was MM-1, our predictor variable was condition; we

652 allowed intercepts and the effect of condition to vary across participants. Overall, the

653 model significantly predicted participants responses, and provided a better fit for the data

654 than the baseline model, $\chi^2(3) = 8.67, p = .034$. Condition significantly influenced MM-1

655 responses $F(1, 872.00) = 7.01, p = .008$, and was a significant predictor in the model $b =$

656 $-0.69, t(872.00) = -2.65, p = .008$, see Figure 12.

657 Differences in the *Good* Descriptions

658 For the *good* characters, we conducted a linear-mixed-effects model to test if

659 condition influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor

660 variable was condition; we allowed intercepts and the effect of condition to vary across

661 participants. Overall, the model significantly predicted participants responses, and

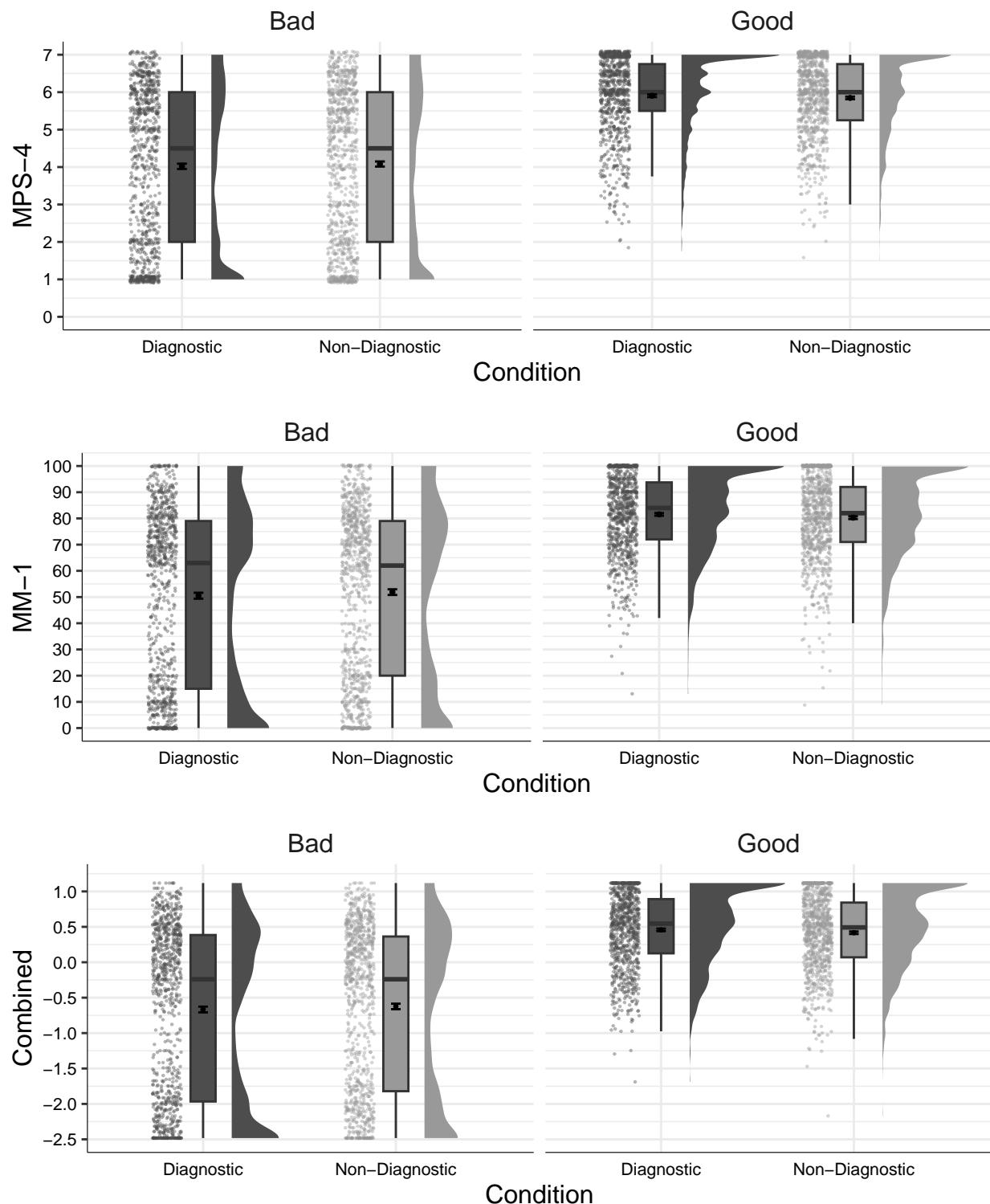


Figure 12. Study S2: Differences in moral perception depending on condition

662 provided a better fit for the data than the baseline model, $\chi^2(3) = 13.66, p = .003$.
663 Condition significantly influenced MPS-4 responses $F(1, 872.00) = 6.82, p = .009$, and was
664 a significant predictor in the model $b = 0.03, t(872.00) = 2.61, p = .009$, see Figure 12.

665 We conducted a linear-mixed-effects model to test if condition influenced MM-1
666 responses. Our outcome measure was MM-1, our predictor variable was condition; we
667 allowed intercepts and the effect of condition to vary across participants. Overall, the
668 model significantly predicted participants responses, and provided a better fit for the data
669 than the baseline model, $\chi^2(1) = 11.97, p < .001$. Condition significantly influenced MM-1
670 responses $F(1, 873) = 12.04, p < .001$, and was a significant predictor in the model $b =$
671 $0.63, t(873) = 3.47, p < .001$, see Figure 12.

672 Study S2: Differences between the descriptions

673 Again, we conducted separate analyses to investigate of condition on responses to
674 each scenario individually. The responses for each scenario across each measure depending
675 on condition are displayed in Figure 13.

676 For *Sam (good)*, MPS-4 scores were significantly lower in the non-diagnostic
677 condition ($M = 5.81, SD = 1.09$), than in the diagnostic condition ($M = 5.98, SD = 0.97$),
678 $t(859.15) = 2.46, p = .014, d = 0.17$; Similarly, MM-1 ratings were significantly lower in
679 the non-diagnostic condition ($M = 79.64, SD = 15.68$), than in the diagnostic condition
680 ($M = 82.37, SD = 14.67$), $t(867.08) = 2.66, p = .008, d = 0.18$. For the combined measure
681 ratings were also lower in the non-diagnostic condition ($M = 0.39, SD = 0.54$), than in the
682 diagnostic condition ($M = 0.50, SD = 0.50$), $t(863.14) = 2.85, p = .004, d = 0.19$.

683 For *Robin (good)*, MPS-4 scores were not significantly different for the non-diagnostic
684 condition ($M = 5.88, SD = 0.96$), than in the diagnostic condition ($M = 5.83, SD = 1.14$),
685 $t(844.53) = -0.77, p = .440, d = 0.05$; MM-1 ratings were similar in the non-diagnostic
686 condition ($M = 80.92, SD = 15.27$), and in the diagnostic condition ($M = 80.70, SD =$

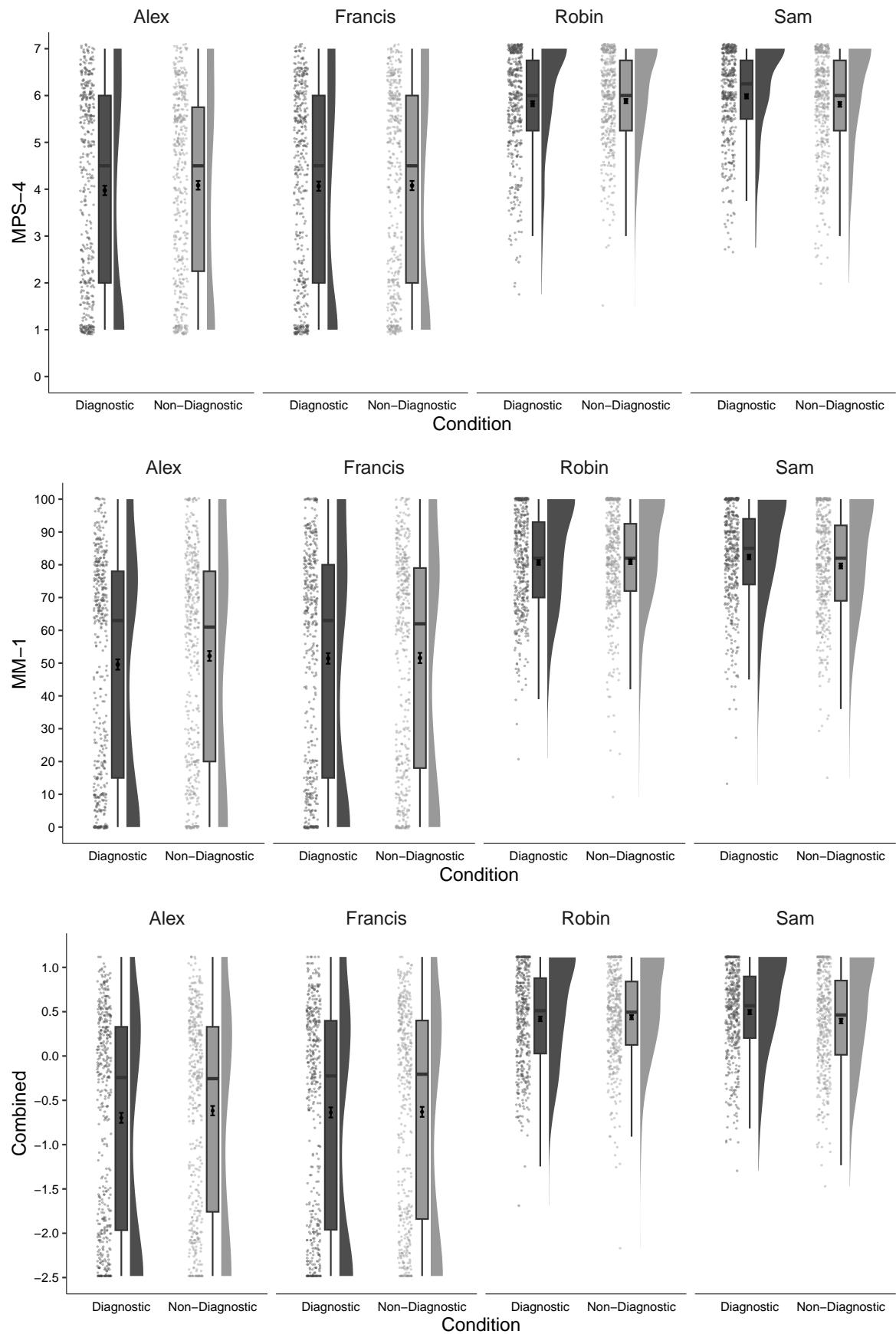


Figure 13. Study 3: Differences in moral perception for each description

687 15.07), $t(871.98) = -0.22$, $p = .828$, $d = 0.01$. For the combined measure ratings were also
688 similar in the non-diagnostic condition ($M = 0.44$, $SD = 0.51$), than in the diagnostic
689 condition ($M = 0.42$, $SD = 0.54$), $t(867.63) = -0.57$, $p = .569$, $d = 0.04$.

690 For *Alex (bad)*, MPS-4 scores were not significantly different for the non-diagnostic
691 condition ($M = 4.08$, $SD = 1.96$), than in the diagnostic condition ($M = 3.97$, $SD = 2.11$),
692 $t(865.81) = -0.80$, $p = .421$, $d = 0.05$; MM-1 ratings were similar in the non-diagnostic
693 condition ($M = 52.19$, $SD = 31.29$), and in the diagnostic condition ($M = 49.58$, $SD =$
694 32.95), $t(868.76) = -1.20$, $p = .230$, $d = 0.08$. For the combined measure ratings were also
695 similar in the non-diagnostic condition ($M = -0.62$, $SD = 1.11$), and in the diagnostic
696 condition ($M = -0.70$, $SD = 1.19$), $t(867.67) = -1.04$, $p = .301$, $d = 0.07$.

697 For *Francis (bad)*, MPS-4 scores were not significantly different for the non-diagnostic
698 condition ($M = 4.08$, $SD = 2.07$), than in the diagnostic condition ($M = 4.07$, $SD = 2.07$),
699 $t(871.94) = -0.09$, $p = .928$, $d = 0.01$; MM-1 ratings were not significantly different in the
700 non-diagnostic condition ($M = 51.56$, $SD = 32.68$), than in the diagnostic condition ($M =$
701 51.42, $SD = 33.70$), $t(871.59) = -0.06$, $p = .952$, $d = 0.00$. For the combined measure
702 ratings were also similar in the non-diagnostic condition ($M = -0.63$, $SD = 1.18$), and in
703 the diagnostic condition ($M = -0.64$, $SD = 1.20$), $t(871.88) = -0.08$, $p = .939$, $d = 0.01$.

Study S3 - Good and Bad Characters

The aim of Study S3 was to test for the moral dilution effect in both good and bad characters, while attempting to eliminate the confounding influence of the presence of other descriptions by adopting a between-subjects design.

Study S3: Method

Study S3: Participants and design. Study S3 was a 2×2 between-subjects factorial design. As in Study 3, 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 (MPS-4, $\alpha = 0.97$, and MM-1).

A total sample of 2389 (1137 female, 1236 male, 5 non-binary, 3 other; 8 prefer not to say, $M_{age} = 38.78$, min = 2, max = 1995, $SD = 42.71$) started the survey. Participants were recruited from MTurk and paid \$0.10 for their participation.

Participants who failed both manipulation checks were removed ($n = 445$), leaving a total sample of 1944 participants (960 female, 970 male, 2 other, 2 prefer not to say; $M_{age} = 37.88$, min = 2, max = 454, $SD = 15.49$).

Study S3: Procedure and materials. The materials for Study S3 were the same as those used in Study 3. Participants were randomly presented with a single character description: *Sam, Robin* (good characters), *Francis* and *Alex* (bad characters), and were randomly assigned to the diagnostic condition (containing diagnostic information only), or the non-diagnostic condition (where the character description additionally included non-diagnostic information). Study S3 was not pre-registered however our predictions were the same as those for Study 3.

727 **Study S3: Results**

728 The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*
729 (good), $M_{MPS-4} = 6.15$, $SD_{MPS-4} = 0.87$, *Francis* (bad), $M_{MPS-4} = 3.65$, $SD_{MPS-4} = 2.16$,
730 *Alex* (bad), $M_{MPS-4} = 3.65$, $SD_{MPS-4} = 2.09$, *Robin* (good), $M_{MPS-4} = 6.21$, $SD_{MPS-4} =$
731 0.85. There was significant variation depending on the description, $F(3,1940) = 396.86$, p
732 $< .001$, partial $\eta^2 = 0.38$. Both the *good* characters (*Robin* and *Sam*) were rated
733 significantly more favorably than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$).
734 There were no differences between *Robin* and *Sam* (*good*: $p = .932$) or between *Alex* and
735 *Francis* (*bad*; $p > .999$).

736 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
737 (good), $M_{MM-1} = 84.70$, $SD_{MM-1} = 15.32$; *Francis* (bad), $M_{MM-1} = 43.37$, $SD_{MM-1} = 34.96$;
738 *Alex* (bad), $M_{MM-1} = 44.68$, $SD_{MM-1} = 34.57$; *Robin* (good), $M_{MM-1} = 85.33$, $SD_{MM-1} =$
739 14.47. There was significant variation depending on the description, $F(3,1940) = 383.99$, p
740 $< .001$, partial $\eta^2 = 0.37$. Both the *good* characters (*Robin* and *Sam*) were rated
741 significantly more favorably than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$).
742 There were no differences between *Robin* and *Sam* (*good*: $p = .982$) or between *Alex* and
743 *Francis* (*bad*; ($p = .872$)).

744 We conducted a 2×2 between subjects ANOVA to test for an interaction between
745 valence and condition in predicting MPS-4. Condition significantly influenced responses to
746 the MPS-4, $F(1, 1940) = 5.16$, $p = .023$; valence significantly predicted responses, $F(1,$
747 $1940) = 1,495.09$, $p < .001$; and there was no significant condition \times valence interaction,
748 $F(1, 1940) = 0.03$, $p = .858$.

749 We conducted a 2×2 between subjects ANOVA to test for an interaction between
750 valence and condition in predicting responses to MM-1. Condition significantly influenced
751 responses to MM-1, $F(1, 1940) = 9.46$, $p = .002$; valence significantly predicted responses,
752 $F(1, 1940) = 580.03$, $p < .001$; and there was no significant condition \times valence

753 interaction, $F(1, 1940) = 0.32, p = .573.$

754 We conducted a 2×2 between subjects ANOVA to test for an interaction between
755 valence and condition in predicting responses to the combined measure. Condition
756 significantly influenced responses to the combined measure, $F(1, 1940) = 9.46, p = .002;$
757 valence significantly predicted responses, $F(1, 1940) = 580.03, p < .001$; and there was no
758 significant condition \times valence interaction, $F(1, 1940) = 0.32, p = .573.$

759 As in previous studied we conducted separate analyses for the good and bad
760 descriptions.

761 Differences in the *Bad* Descriptions

762 For the *bad* characters, there was no significant difference in responses to MPS-4
763 between the diagnostic condition ($M = 3.57, SD = 2.21$) and the non-diagnostic condition
764 ($M = 3.72, SD = 2.03$) depending on condition, $t(941.99) = -1.09, p = .277, d = 0.07.$

765 For the *bad* characters, there was no significant difference in responses to MM-1
766 between the diagnostic condition ($M = 34.62, SD = 34.25$) and the non-diagnostic
767 condition ($M = 37.04, SD = 32.16$) depending on condition, $t(680.45) = -0.96, p = .339, d$
768 = 0.07.

769 Differences in the *Good* Descriptions

770 For the *good* characters, there was no significant difference in responses to MPS-4
771 between the diagnostic condition ($M = 6.27, SD = 0.84$) and the non-diagnostic condition
772 ($M = 6.09, SD = 0.88$) depending on condition, $t(981.90) = 3.21, p = .001, d = 0.20.$

773 For the *good* characters, there was no significant difference in responses to MPS-4
774 between the diagnostic condition ($M = 87.23, SD = 13.68$) and the non-diagnostic
775 condition ($M = 82.89, SD = 15.70$) depending on condition, $t(972.88) = 4.63, p < .001, d$
776 = 0.29.

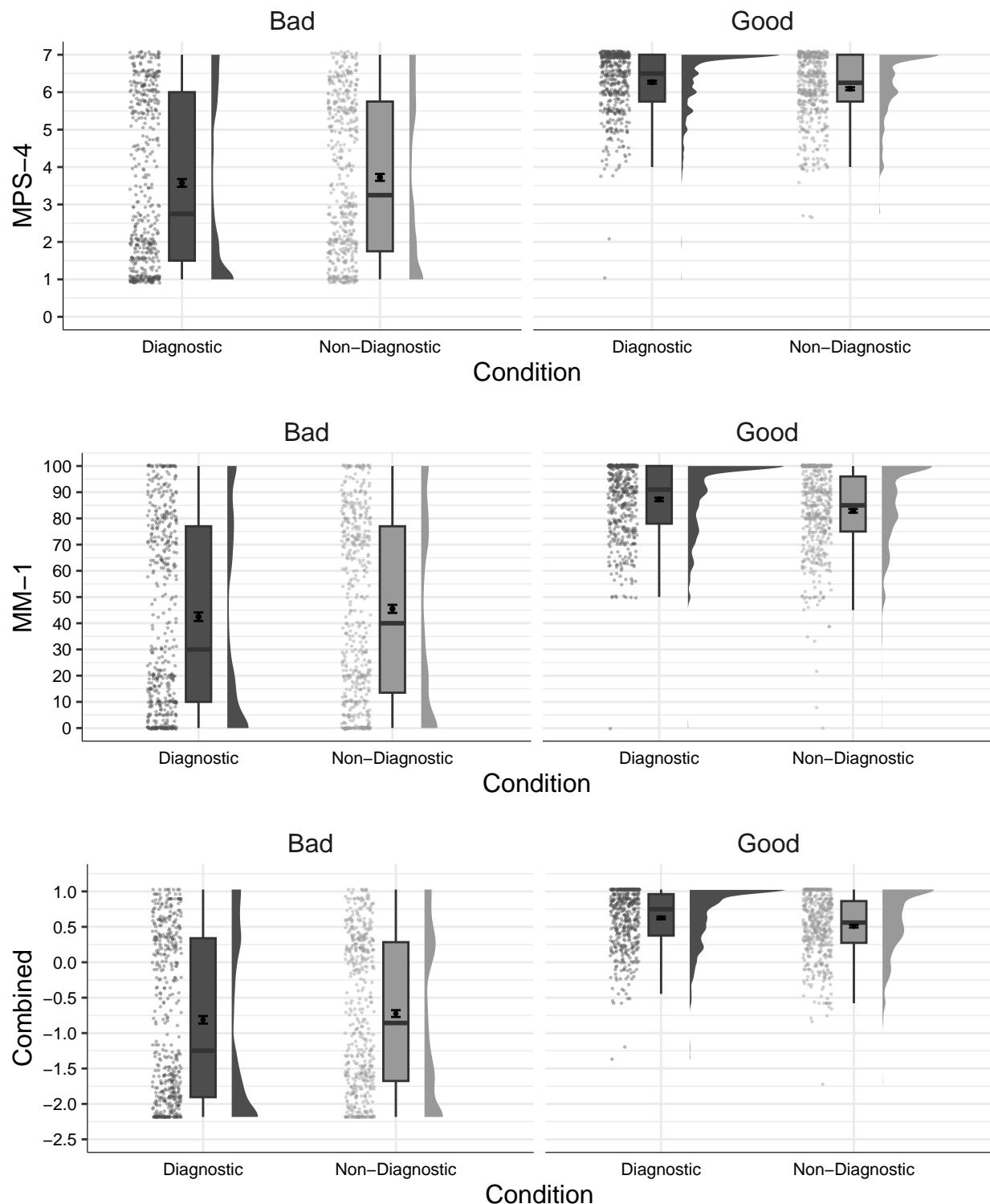


Figure 14. Study S3: Differences in moral perception depending on condition

777 Study S3: Differences between the descriptions

778 Again, we conducted separate analyses to investigate of condition on responses to
779 each scenario individually. The responses for each scenario across each measure depending
780 on condition are displayed in Figure 15.

781 For *Sam*, MPS-4 scores were significantly lower in the non-diagnostic condition ($M =$
782 6.06 , $SD = 0.89$), than in the diagnostic condition ($M = 6.24$, $SD = 0.85$), $t(488.52) =$
783 2.34 , $p = .020$, $d = 0.21$; Similarly, MM-1 ratings were significantly lower in the
784 non-diagnostic condition ($M = 82.23$, $SD = 16.63$), than in the diagnostic condition ($M =$
785 87.21 , $SD = 13.43$), $t(471.87) = 3.65$, $p < .001$, $d = 0.33$. For the combined measure
786 ratings were also lower in the non-diagnostic condition ($M = 0.49$, $SD = 0.45$), than in the
787 diagnostic condition ($M = 0.62$, $SD = 0.41$), $t(486.39) = 3.35$, $p < .001$, $d = 0.30$.

788 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
789 condition ($M = 6.13$, $SD = 0.87$), than in the diagnostic condition ($M = 6.30$, $SD = 0.83$),
790 $t(490.93) = 2.21$, $p = .027$, $d = 0.20$; MM-1 ratings were similar in the non-diagnostic
791 condition ($M = 83.53$, $SD = 14.75$), and in the diagnostic condition ($M = 87.25$, $SD =$
792 13.96), $t(490.98) = 2.88$, $p = .004$, $d = 0.26$. For the combined measure ratings were also
793 similar in the non-diagnostic condition ($M = 0.53$, $SD = 0.42$), than in the diagnostic
794 condition ($M = 0.63$, $SD = 0.40$), $t(490.92) = 2.82$, $p = .005$, $d = 0.25$.

795 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
796 condition ($M = 3.71$, $SD = 2.02$), than in the diagnostic condition ($M = 3.59$, $SD = 2.16$),
797 $t(465.98) = -0.62$, $p = .534$, $d = 0.06$; MM-1 ratings were similar in the non-diagnostic
798 condition ($M = 45.80$, $SD = 33.76$), than in the diagnostic condition ($M = 43.48$, $SD =$
799 35.46), $t(468.21) = -0.73$, $p = .465$, $d = 0.07$. For the combined measure ratings were also
800 similar in the non-diagnostic condition ($M = -0.72$, $SD = 1.05$), than in the diagnostic
801 condition ($M = -0.79$, $SD = 1.13$), $t(465.61) = -0.69$, $p = .489$, $d = 0.06$.

802 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic

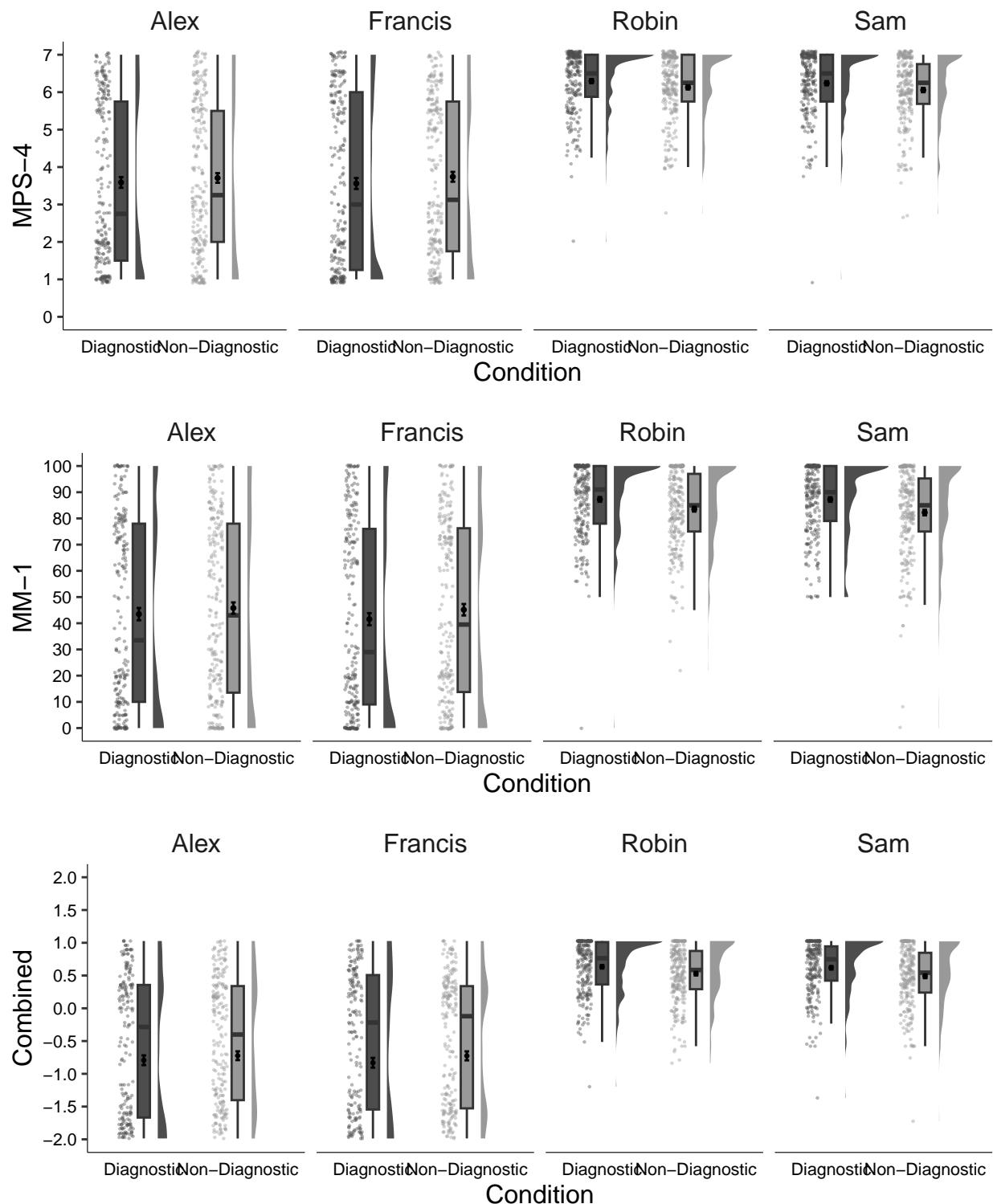


Figure 15. Study S3: Differences in moral perception for each description

803 condition ($M = 3.74$, $SD = 2.05$), than in the diagnostic condition ($M = 3.56$, $SD = 2.27$),
804 $t(474.15) = -0.91$, $p = .364$, $d = 0.08$; MM-1 ratings were not significantly different in the
805 non-diagnostic condition ($M = 45.16$, $SD = 34.18$), than in the diagnostic condition ($M =$
806 41.55 , $SD = 35.73$), $t(478.97) = -1.13$, $p = .258$, $d = 0.10$. For the combined measure
807 ratings were also similar in the non-diagnostic condition ($M = -0.73$, $SD = 1.07$), and in
808 the diagnostic condition ($M = -0.83$, $SD = 1.16$), $t(476.42) = -1.04$, $p = .297$, $d = 0.10$.

809

References

- 810 Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., ... Eden,
811 A. (2020). Validating the extended character morality questionnaire. *Media Psychology*,
812 23(1), 107–130. <https://doi.org/10.1080/15213269.2019.1572523>
- 813 Walker, A. C., Turpin, M. H., Fugelsang, J. A., & Białek, M. (2021). Better the two devils
814 you know, than the one you don't: Predictability influences moral judgments of
815 immoral actors. *Journal of Experimental Social Psychology*, 97, 104220.
816 <https://doi.org/10.1016/j.jesp.2021.104220>