

¹ Supplement: Moral Dilution

² Cillian McHugh¹ & Eric R. Igou¹

³ ¹ University of Limerick

⁴ Author Note

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

18 *Keywords:* keywords

19 Word count: X

20 Supplement: Moral Dilution

21 **Contents**

22 Supplementary 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

Supplementary 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 known
40 to physically hurt others, treat some people differently to others, and show lack of loyalty.

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

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

46 **Non-Diagnostic Descriptions.**

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

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

51 **Descriptions (Good Characters)**

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

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

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

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

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

67 **Non-Diagnostic**

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

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

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

73 **Diagnostic Descriptions.**

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

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

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

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

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

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

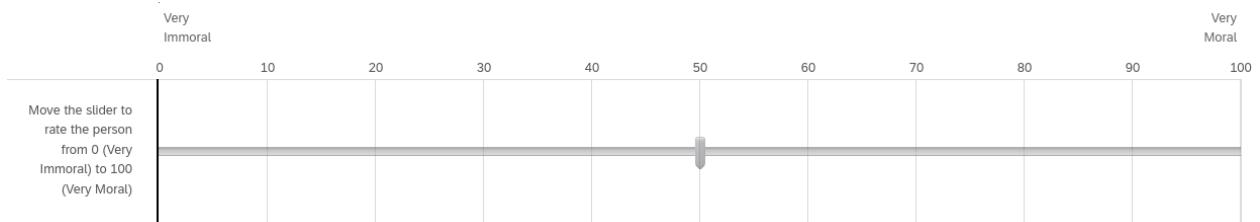
⁸⁹ **Measures**

⁹⁰ **Four-item Moral Perception Scale (MPS-4).** Please rate _____ along the
⁹¹ 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

⁹² **Single-item Moral Perception Measure (MM-1).** Please rate _____ according
⁹³ to immoral or moral you view them:



⁹⁴

95

Study 1 (bad): Supplementary Analyses

96 Study 1: Combined Measure

97 We developed a combined moral perception measure by calculating the mean of the
 98 combined mean-centered scores for MPS-4 and MM-1, and mean-centering this result. Below
 99 we report the analyses for this combined measure.

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

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

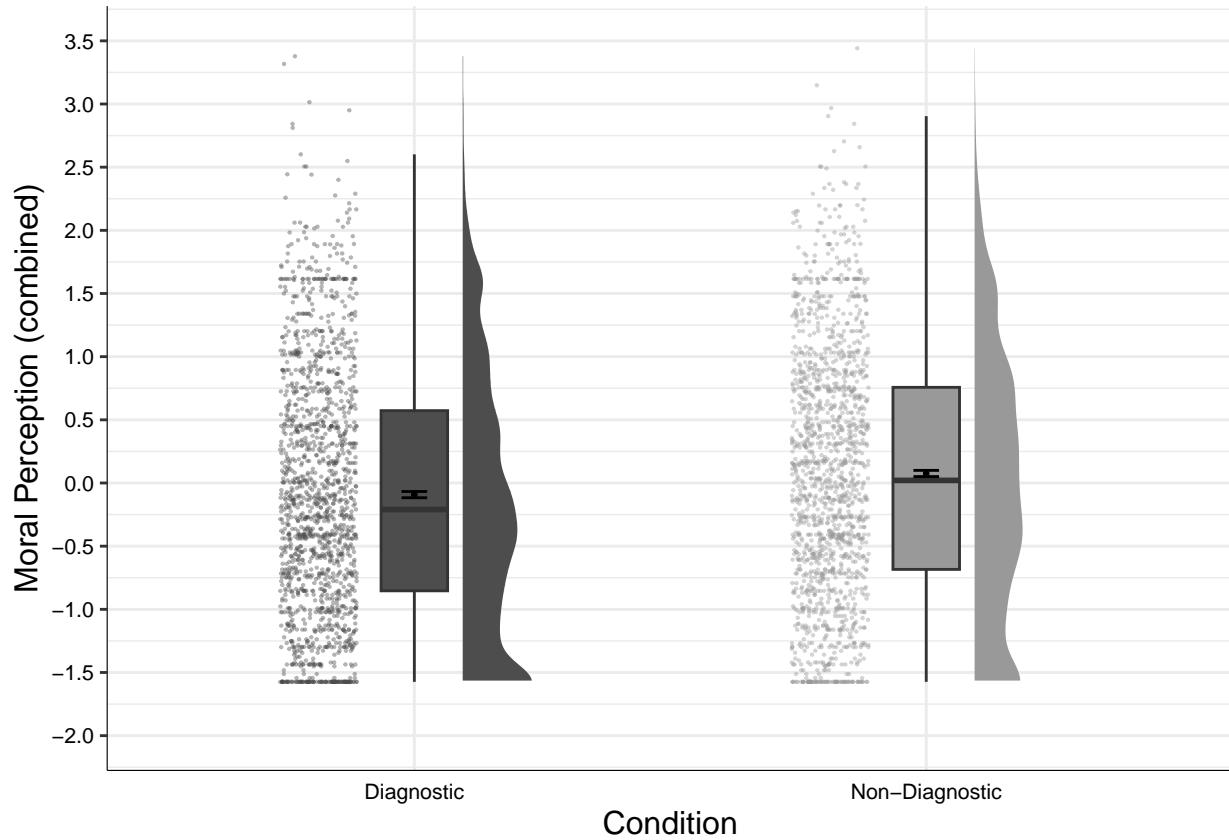


Figure 2. Study 1: Differences in combined measure depending on condition

¹¹⁶ **Study 1: Differences between the Descriptions**

¹¹⁷ We additionally conducted separate analyses for each scenario individually (for each
¹¹⁸ dependent measure MPS-4, MM-1 and the combined measure). The responses for each
¹¹⁹ scenario across each measure depending on condition are displayed in Figure 3.

¹²⁰ For *Sam*, MPS-4 scores were significantly higher for the non-diagnostic condition ($M =$
¹²¹ 2.70 , $SD = 0.82$), than in the diagnostic condition ($M = 2.42$, $SD = 0.87$), $t(798.90) = -4.66$,
¹²² $p < .001$, $d = 0.33$; MM-1 ratings were higher in the non-diagnostic condition ($M = 26.55$,
¹²³ $SD = 16.41$), than in the diagnostic condition ($M = 21.50$, $SD = 15.59$), $t(787.84) = -4.45$, p
¹²⁴ $< .001$, $d = 0.32$. For the combined measure ratings were also higher in the non-diagnostic
¹²⁵ condition ($M = 0.18$, $SD = 0.88$), than in the diagnostic condition ($M = -0.13$, $SD = 0.88$),
¹²⁶ $t(795.41) = -4.98$, $p < .001$, $d = 0.35$.

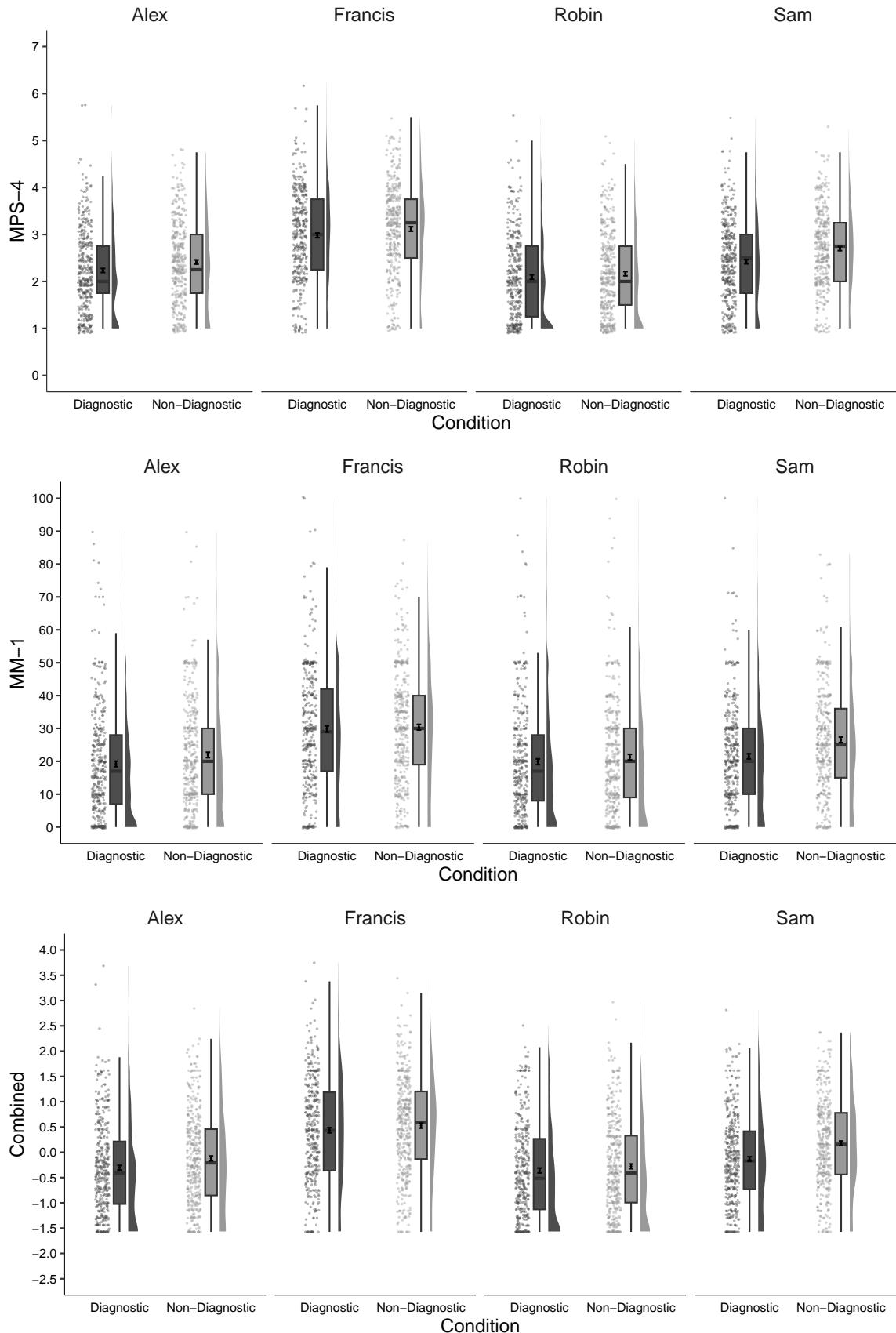


Figure 3. Study 1: Differences in moral perception for each description

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

134 For *Alex*, MPS-4 scores were significantly higher for the non-diagnostic condition ($M =$
135 2.41 , $SD = 0.88$), than in the diagnostic condition ($M = 2.23$, $SD = 0.86$), $t(796.97) = -2.92$,
136 $p = .004$, $d = 0.21$; MM-1 ratings were higher in the non-diagnostic condition ($M = 21.93$,
137 $SD = 16.47$), than in the diagnostic condition ($M = 19.20$, $SD = 16.73$), $t(798.89) = -2.33$, p
138 $= .020$, $d = 0.16$. For the combined measure ratings were also higher in the non-diagnostic
139 condition ($M = -0.12$, $SD = 0.92$), than in the diagnostic condition ($M = -0.30$, $SD = 0.92$),
140 $t(798.40) = -2.82$, $p = .005$, $d = 0.20$.

141 For *Francis*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
142 $= 3.12$, $SD = 0.95$), than in the diagnostic condition ($M = 2.98$, $SD = 0.97$), $t(796.12) =$
143 -1.99 , $p = .047$, $d = 0.14$; MM-1 ratings were not significantly different in the non-diagnostic
144 condition ($M = 30.38$, $SD = 17.17$), than in the diagnostic condition ($M = 29.84$, $SD =$
145 18.56), $t(788.61) = -0.43$, $p = .668$, $d = 0.03$. For the combined measure ratings were also
146 similar in the non-diagnostic condition ($M = 0.53$, $SD = 0.98$), and in the diagnostic
147 condition ($M = 0.44$, $SD = 1.02$), $t(794.36) = -1.29$, $p = .198$, $d = 0.09$.

148

Study 2 (good): Supplementary Analyses

149 **Study 2: Combined Measure**

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

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

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

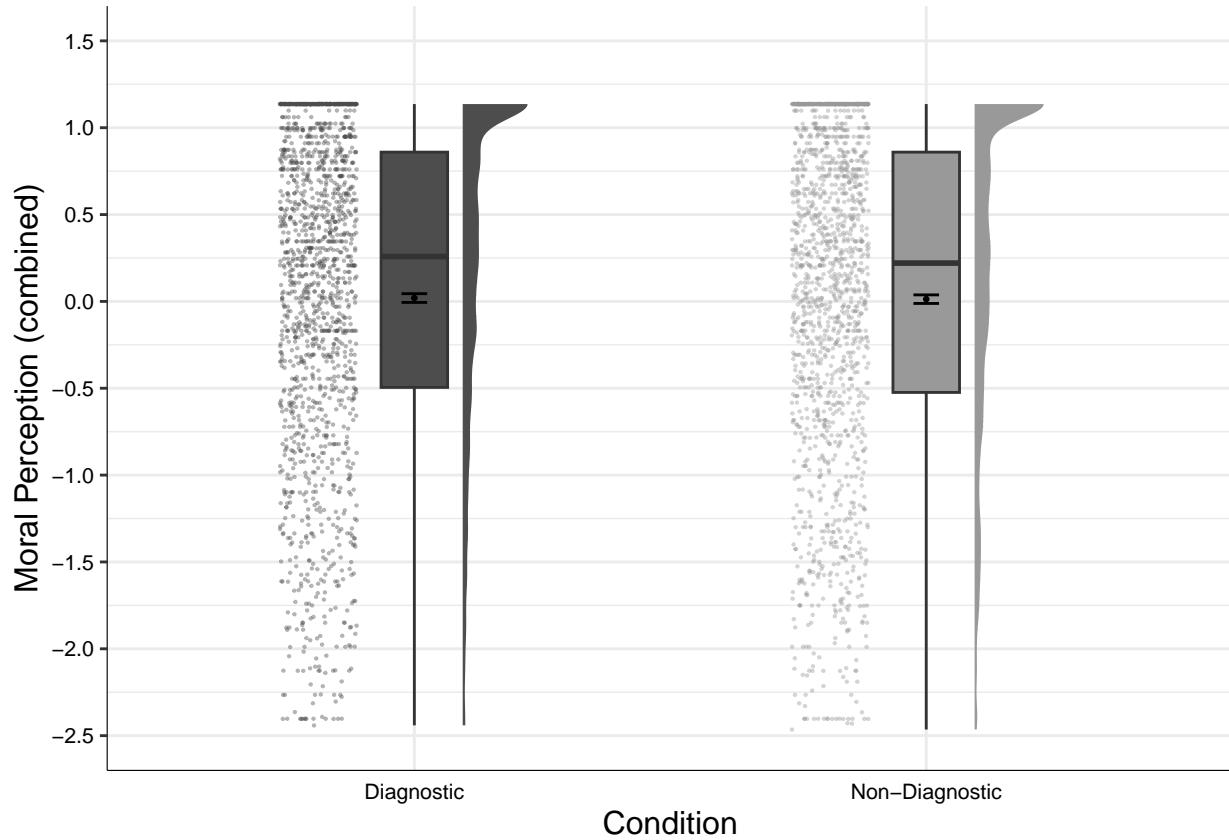


Figure 4. Study 2: Differences in combined measure depending on condition

¹⁶⁶ **Study 2: Differences between the Descriptions**

¹⁶⁷ Below we provide analyses of the effect of condition on responses to each scenario
¹⁶⁸ individually. The responses for each scenario across each measure depending on condition are
¹⁶⁹ displayed in Figure 5.

¹⁷⁰ For *Sam*, MPS-4 scores were not significantly different in the non-diagnostic condition
¹⁷¹ ($M = 6.17$, $SD = 0.89$), than in the diagnostic condition ($M = 6.05$, $SD = 1.06$), $t(680.49) =$
¹⁷² -1.71 , $p = .088$, $d = 0.12$; MM-1 ratings were similar in the non-diagnostic condition ($M =$
¹⁷³ 84.90 , $SD = 14.26$), and in the diagnostic condition ($M = 84.20$, $SD = 14.76$), $t(744.17) =$
¹⁷⁴ -0.69 , $p = .490$, $d = 0.05$. For the combined measure ratings were also similar in the
¹⁷⁵ non-diagnostic condition ($M = 0.11$, $SD = 0.93$), and in the diagnostic condition ($M = 0.02$,
¹⁷⁶ $SD = 1.03$), $t(717.94) = -1.33$, $p = .183$, $d = 0.10$.

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

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

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

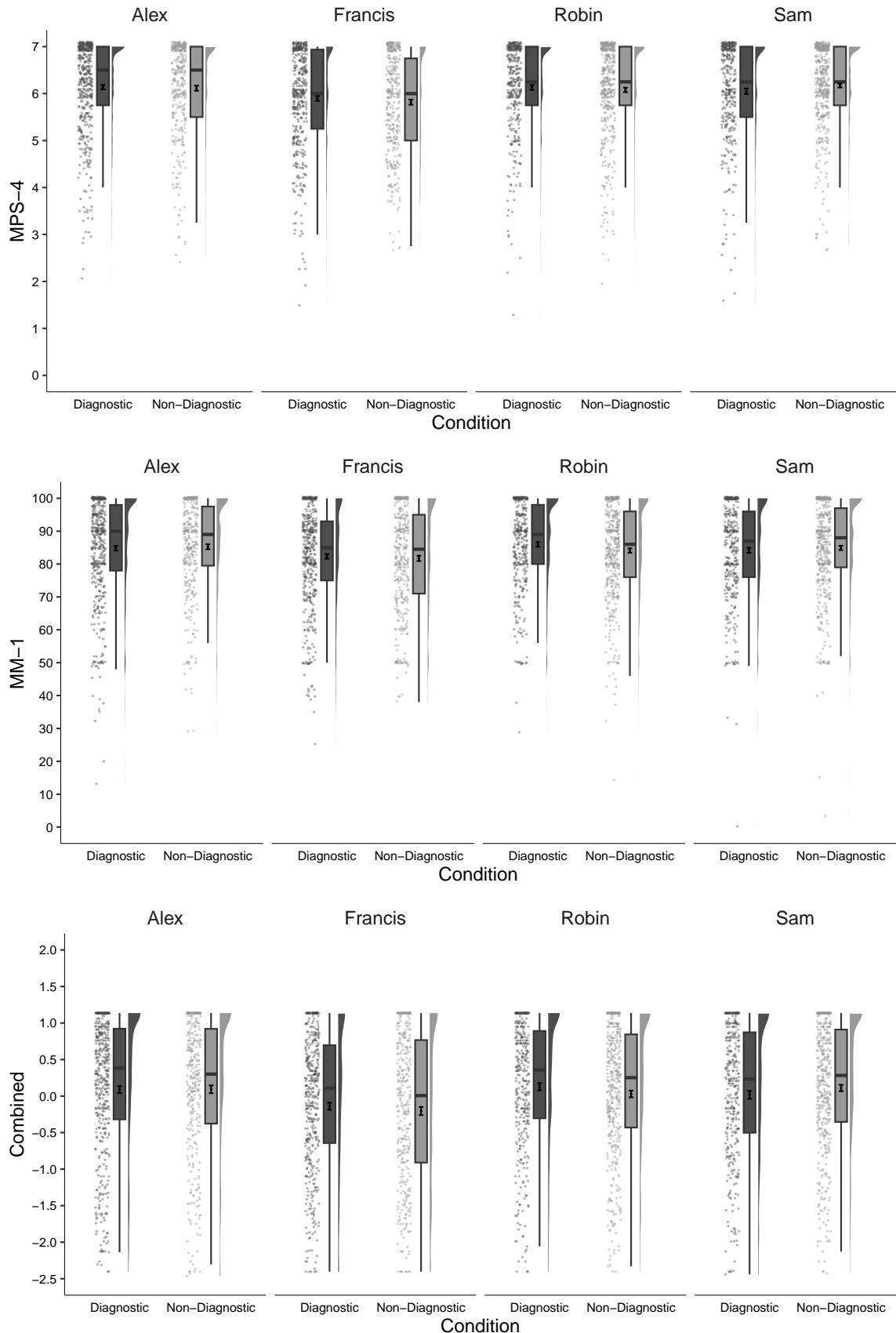


Figure 5. Study 2: Differences in moral perception for each description

198 Study 3 (bad and good): Supplementary Analyses**199 Study 3: Combined Measure**

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

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

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

220 For the *bad* characters, we conducted a linear-mixed-effects model to test if condition
221 influenced responses to the combined measure. Our outcome measure was the combined
222 moral perception measure, 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, $\chi^2(3) = 74.54, p < .001$. Condition significantly influenced MPS-4 responses $F(1, 820.39) = 37.63, p < .001$, and was a significant predictor in the model $b = -0.04, t(820.39) = -6.13, p < .001$.

For the *good* characters, we conducted a linear-mixed-effects model to test if condition influenced responses to the combined measure. Our outcome measure was the combined moral perception measure, 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, $\chi^2(3) = 45.20, p < .001$. Condition significantly influenced MPS-4 responses $F(1, 826.21) = 15.67, p < .001$, and was a significant predictor in the model $b = 0.02, t(826.21) = 3.96, p < .001$.

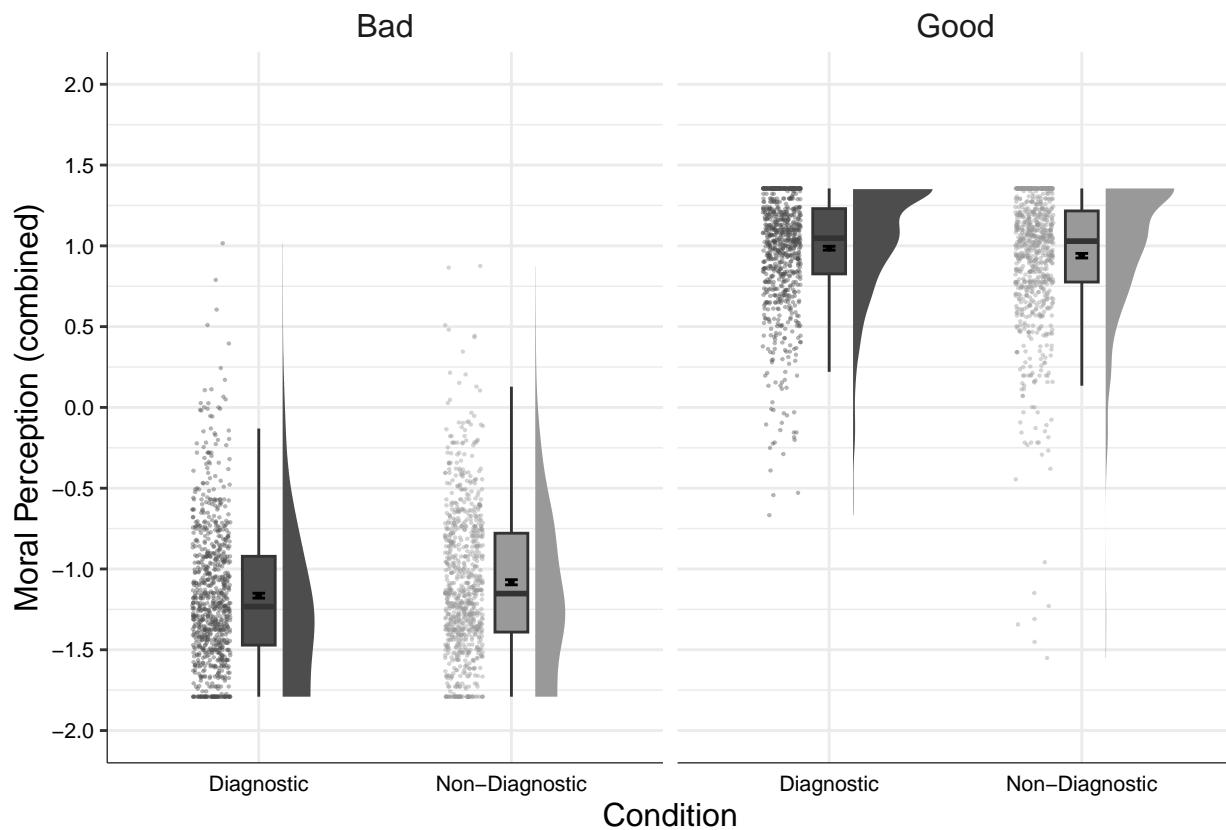


Figure 6. Study 3: Differences in the combined measure depending on condition

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

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

238 For *Sam*, MPS-4 scores were not significantly lower in the non-diagnostic condition (M
239 = 6.15, SD = 0.86), than in the diagnostic condition (M = 6.25, SD = 0.76), $t(812.83)$ =
240 1.68, p = .094, d = 0.12; Similarly, MM-1 ratings were similar in the non-diagnostic
241 condition (M = 85.49, SD = 14.10), in the diagnostic condition (M = 87.18, SD = 13.21),
242 $t(821.76)$ = 1.78, p = .075, d = 0.12. For the combined measure ratings was also no
243 significant difference between the non-diagnostic condition (M = 0.90, SD = 0.42), and the
244 diagnostic condition (M = 0.96, SD = 0.37), $t(811.12)$ = 1.88, p = .060, d = 0.13.

245 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
246 condition (M = 6.28, SD = 0.80), than in the diagnostic condition (M = 6.36, SD = 0.71),
247 $t(809.44)$ = 1.60, p = .111, d = 0.11; MM-1 ratings were similar in the non-diagnostic
248 condition (M = 87.84, SD = 13.49), and in the diagnostic condition (M = 89.02, SD =
249 10.30), $t(765.30)$ = 1.42, p = .156, d = 0.10. For the combined measure ratings were also
250 similar in the non-diagnostic condition (M = 0.97, SD = 0.39), than in the diagnostic
251 condition (M = 1.01, SD = 0.32), $t(784.03)$ = 1.63, p = .103, d = 0.11.

252 For *Alex*, MPS-4 scores were significantly higher for the non-diagnostic condition (M =
253 2.41, SD = 0.88), than in the diagnostic condition (M = 2.24, SD = 0.90), $t(830.38)$ = -2.69,
254 p = .007, d = 0.19; MM-1 ratings were similar in the non-diagnostic condition (M = 23.53,
255 SD = 16.61), and in the diagnostic condition (M = 22.62, SD = 18.34), $t(828.19)$ = -0.75, p
256 = .454, d = 0.05. For the combined measure ratings were also similar in the non-diagnostic
257 condition (M = -1.05, SD = 0.45), and in the diagnostic condition (M = -1.11, SD = 0.47),
258 $t(830.90)$ = -1.77, p = .077, d = 0.12.

259 For *Francis*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
260 = 2.26, SD = 0.85), than in the diagnostic condition (M = 2.05, SD = 0.70), $t(802.80)$ =
261 -3.96, p < .001, d = 0.27; MM-1 ratings were significantly higher in the non-diagnostic
262 condition (M = 22.01, SD = 17.84), than in the diagnostic condition (M = 18.45, SD =
263 15.76), $t(817.94)$ = -3.05, p = .002, d = 0.21. For the combined measure ratings were also
264 significantly higher in the non-diagnostic condition (M = -1.11, SD = 0.46), than in the
265 diagnostic condition (M = -1.23, SD = 0.38), $t(808.55)$ = -3.85, p < .001, d = 0.27.

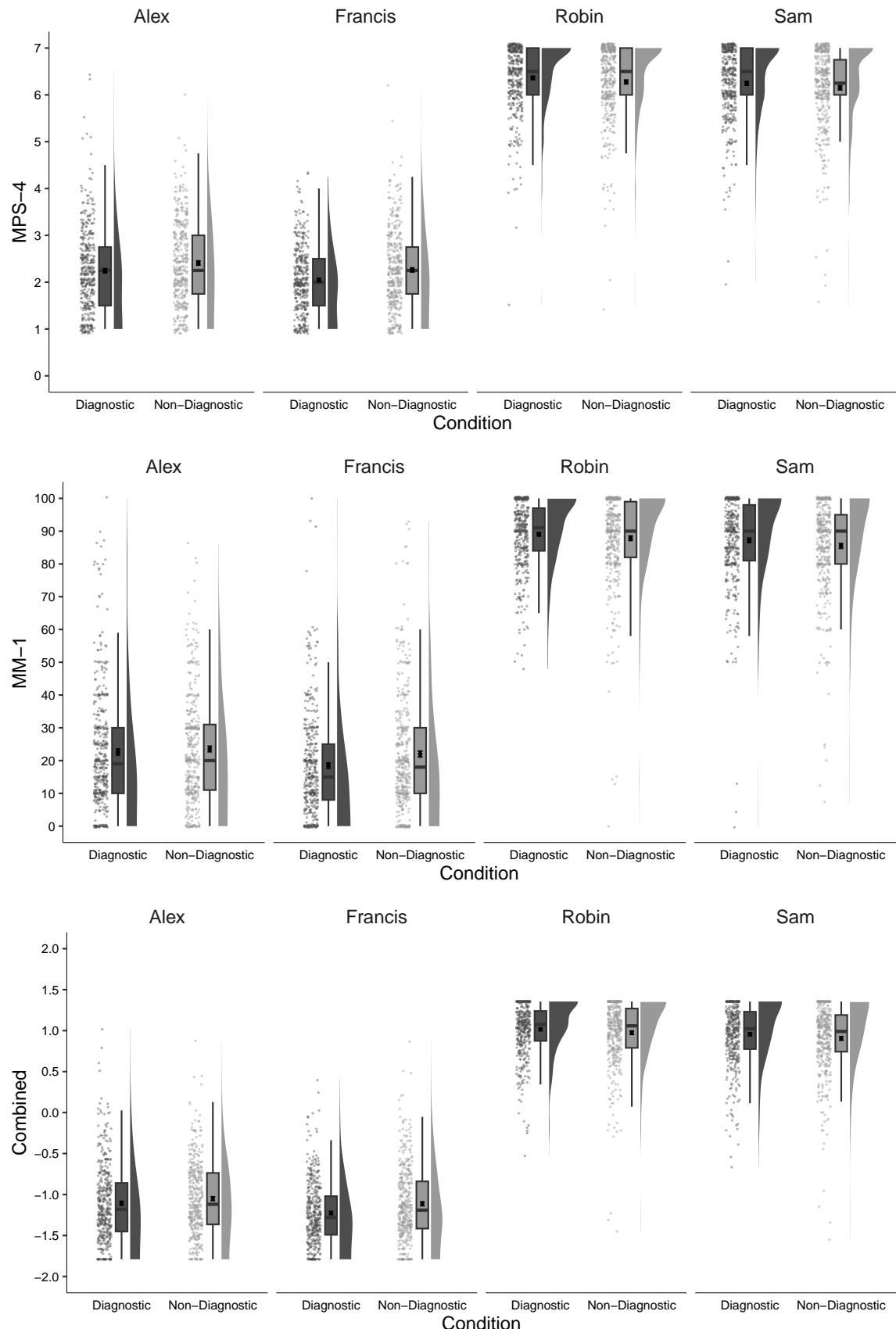


Figure 7. Study 3: Differences in moral perception for each description

266

Pilot Study 1

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

271 **Pilot Study 1: Method**

272 **Pilot 1: Participants and design.** The pilot study was a within-subjects design.
273 The independent variable was description type with two levels, *diagnostic* and
274 *non-diagnostic*. We used two dependent variables. The first dependent variable was the four
275 item moral perception scale (MPS-4), participants rated the characters on four dimensions
276 using 7-point bipolar scales. The dimensions and scale endpoints were: Bad-Good,
277 Immoral-Moral, Violent-Peaceful, Merciless-Empathetic, this showed excellent reliability, α
278 = 0.93. The second dependent variable was a single item moral perception measure (MM-1)
279 which consisted of a 100-point slider ranging from 0 = *Very Immoral* to 100 = *Very Moral*.
280 Both dependent variables were taken from Walker et al. (2021).

281 A total sample of 235 (89 female, 142 male, 1 non-binary, 1 prefer not to say; $M_{age} =$
282 36.45, min = 20, max = 72, $SD = 10.23$) started the survey. Participants were recruited
283 from MTurk.

284 We removed participants who failed both manipulation checks ($n = 23$), leaving a total
285 sample of 212 participants (80 female, 128 male, 1 non-binary, 1 prefer not to say; $M_{age} =$
286 36.63, min = 20, max = 72, $SD = 10.34$).

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

290 Moral character descriptions were developed by combining descriptions relating to

291 three different moral foundations. These descriptions were adapted from the items of the
292 extended character morality questionnaire (Grizzard et al., 2020), and read as follows:

- 293 (i) *Imagine a person named Sam. Throughout their life they have been known to be cruel,*
294 *act unfairly, and to betray their own group;*
- 295 (ii) *Imagine a person named Robin. Throughout their life they have been known to*
296 *physically hurt others, treat some people differently to others, and show lack of loyalty;*
- 297 (iii) *Imagine a person named Francis. Throughout their life they have been known to violate*
298 *the standards of purity and decency, show lack of respect for authority, and treat people*
299 *unequally*
- 300 (iv) *Imagine a person named Alex. Throughout their life they have been known to cause*
301 *others to suffer emotionally, to deny others their rights, and to cause chaos or disorder.*

302 We developed neutral descriptions that included information relating to physical
303 appearance/attributes, hobbies/activities, and family information that read as follows:

- 304 (i) *Imagine a person named Jackie. They have red hair, play tennis four times a month,*
305 *and have one older sibling and one younger sibling;*
- 306 (ii) *Imagine a person named Charlie. They are left-handed, drink tea in the morning, and*
307 *have two older siblings and one younger sibling.*

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

312 **Pilot 1: Results**

313 **Pilot 1: Main Measures.** The means and standard deviations for MPS-4 for each
314 scenario are as follows: *Sam* (diagnostic), $M_{MPS-4} = 4.35$, $SD_{MPS-4} = 1.90$, *Francis*

315 (diagnostic), $M_{\text{MPS-4}} = 4.46$, $SD_{\text{MPS-4}} = 1.73$, *Alex* (diagnostic), $M_{\text{MPS-4}} = 4.44$, $SD_{\text{MPS-4}} =$
 316 1.79, *Robin* (diagnostic), $M_{\text{MPS-4}} = 4.35$, $SD_{\text{MPS-4}} = 1.96$, *Jackie* (non-diagnostic), $M_{\text{MPS-4}} =$
 317 5.40, $SD_{\text{MPS-4}} = 1.01$, *Charlie* (non-diagnostic), $M_{\text{MPS-4}} = 5.38$, $SD_{\text{MPS-4}} = 1.01$. For the
 318 diagnostic descriptions, there was no significant variation depending on the description,
 319 $F(3,600) = 1.58$, $p = .194$, partial $\eta^2 = 0.00$. For the non-diagnostic descriptions there was
 320 no significant difference in ratings depending on description, $t(211) = -0.67$, $p = .506$, $d =$
 321 0.05.

322 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
 323 (diagnostic), $M_{\text{MM-1}} = 55.67$, $SD_{\text{MM-1}} = 30.47$; *Francis* (diagnostic), $M_{\text{MM-1}} = 58.22$, $SD_{\text{MM-1}}$
 324 = 28.61; *Alex* (diagnostic), $M_{\text{MM-1}} = 56.80$, $SD_{\text{MM-1}} = 29.45$; *Robin* (diagnostic), $M_{\text{MM-1}} =$
 325 55.49, $SD_{\text{MM-1}} = 31.38$; *Jackie* (non-diagnostic), $M_{\text{MM-1}} = 73.00$, $SD_{\text{MM-1}} = 14.72$; *Charlie*
 326 (non-diagnostic), $M_{\text{MM-1}} = 72.94$, $SD_{\text{MM-1}} = 14.79$. For the diagnostic descriptions, we
 327 observed significant variation depending on the description, $F(3,608) = 3.01$, $p = .032$,
 328 partial $\eta^2 = 0.001$. When correcting for multiple comparisons, pairwise comparisons did not
 329 reveal significant differences between descriptions. We note that without correction, *Francis*
 330 appeared to be rated as more moral than both *Robin* ($p = .012$), and *Sam* ($p = .009$). For
 331 the non-diagnostic descriptions there was no significant difference in ratings depending on
 332 description, $t(211) = -0.09$, $p = .929$, $d = 0.01$.

333 We conducted a linear-mixed-effects model to test if condition influenced MPS-4
 334 responses. Our outcome measure was MPS-4, our predictor variable was condition; we
 335 allowed intercepts and the effect of condition to vary across participants. Overall, the model
 336 significantly predicted participants responses, and provided a better fit for the data than the
 337 baseline model, $\chi^2(2) = 860.16$, $p < .001$. Condition was a significant predictor in the model
 338 $b = -0.49$, $t(211.05) = -8.54$, $p < .001$, with the non-diagnostic descriptions being rated as
 339 more moral than the diagnostic descriptions of immoral characters Figure 8.

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

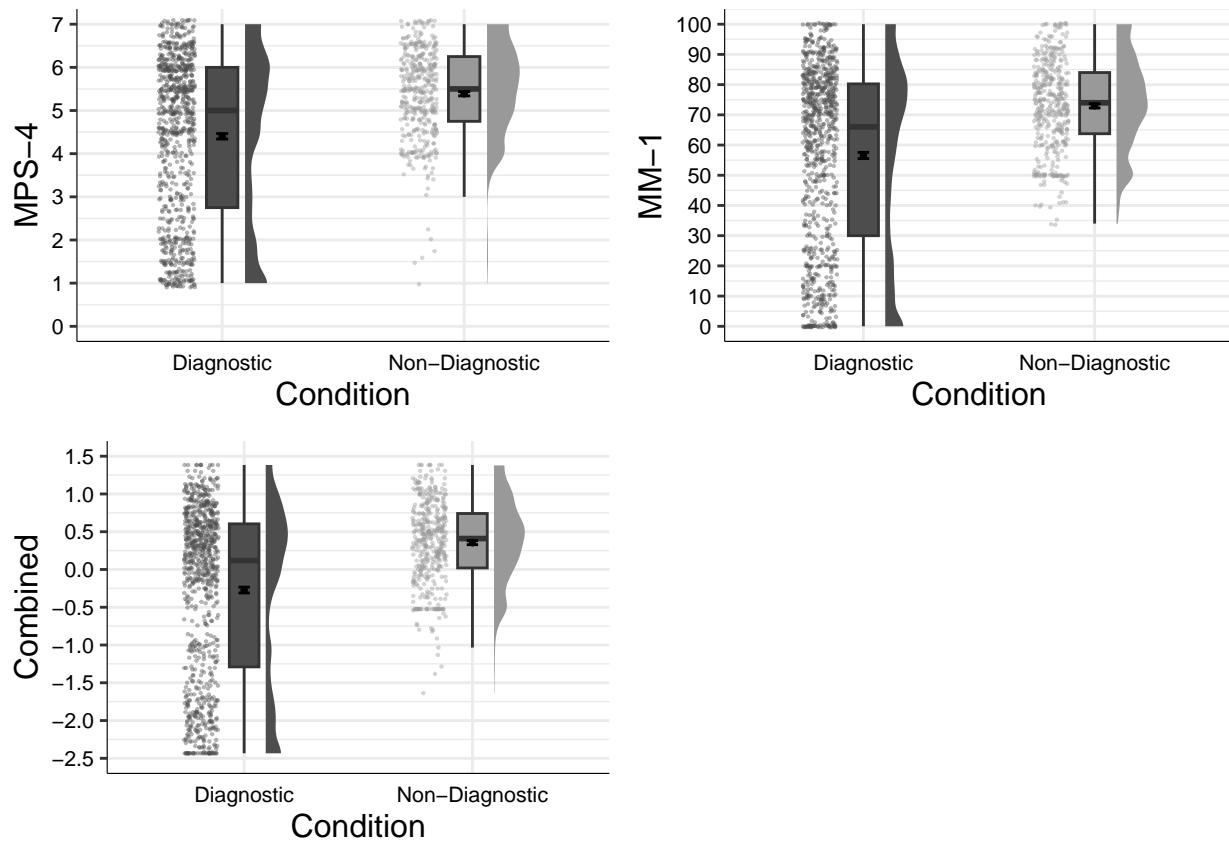


Figure 8. Pilot Study 1: Differences in moral perception depending on condition

341 responses. Our outcome measure was MM-1, our predictor variable was condition; we
 342 allowed intercepts and the effect of condition to vary across participants. Overall, the model
 343 significantly predicted participants responses, and provided a better fit for the data than the
 344 baseline model, $\chi^2(2) = 924.82, p < .001$. Condition was a significant predictor in the model
 345 $b = -8.22, t(210.98) = -8.60, p < .001$, with the non-diagnostic descriptions being rated as
 346 more moral than the diagnostic descriptions, see Figure 8.

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

350 The standardized means and standard deviations for the combined measure for each
 351 scenario are as follows: *Sam* (diagnostic), $M = -0.30, SD = 1.16$; *Francis* (diagnostic), $M =$

352 -0.22, $SD = 1.06$; *Alex* (diagnostic), $M = -0.25$, $SD = 1.10$; *Robin* (diagnostic), $M = -0.31$,
353 $SD = 1.19$; *Jackie* (non-diagnostic), $M = 0.36$, $SD = 0.55$; *Charlie* (non-diagnostic), $M =$
354 0.35, $SD = 0.55$. For the moral descriptions, we observed significant variation depending on
355 the description, $F(3,602) = 2.67$, $p = .050$, partial $\eta^2 = 0.001$. When correcting for multiple
356 comparisons, pairwise comparisons did not reveal significant differences between descriptions.
357 We note that without correction, *Francis* appeared to be rated as more moral than both
358 *Robin* ($p = .022$), and *Sam* ($p = .021$). For the neutral descriptions there was no significant
359 difference in ratings depending on description, $t(211) = -0.46$, $p = .645$, $d = 0.03$.

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

366

Pilot Study 2

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

372 **Pilot Study 2: Method**

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

377 A total sample of 245 (70 female, 175 male, 0 non-binary, 0 prefer not to say; $M_{age} =$
378 36.69, min = 18, max = 71, $SD = 9.57$) started the survey. Participants were recruited from
379 MTurk.

380 We removed participants who failed both manipulation checks ($n = 30$), leaving a total
381 sample of 215 participants (63 female, 152 male, 0 non-binary, 0 prefer not to say; $M_{age} =$
382 36.59, min = 18, max = 71, $SD = 9.59$).

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

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

- 390 (i) *Imagine a person named Sam. Throughout their life they have been known to always*
 391 *help and care for others, treat everyone fairly and equally, and show a strong sense of*
 392 *loyalty to others;*
- 393 (ii) *Imagine a person named Robin. Throughout their life they have been known to show*
 394 *compassion and empathy for others, act with a sense of fairness and justice, and, never*
 395 *to break their word;*
- 396 (iii) *Imagine a person named Francis. Throughout their life they have been known to uphold*
 397 *the standards of purity and decency, show respect for authority, and to always act*
 398 *honestly and fairly;*
- 399 (iv) *Imagine a person named Alex. Throughout their life they have been known to protect*
 400 *and provide shelter to the weak and vulnerable, uphold the rights of others, and show*
 401 *respect for authority.*

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

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

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

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

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

410 Pilot 2: Results

411 **Pilot 2: Main Measures.** The means and standard deviations for MPS-4 for each
 412 scenario are as follows: *Sam* (diagnostic), $M_{MPS-4} = 6.01$, $SD_{MPS-4} = 0.91$, *Francis*
 413 (diagnostic), $M_{MPS-4} = 5.89$, $SD_{MPS-4} = 0.95$, *Alex* (diagnostic), $M_{MPS-4} = 5.94$, $SD_{MPS-4} =$
 414 0.94, *Robin* (diagnostic), $M_{MPS-4} = 5.93$, $SD_{MPS-4} = 0.92$, *Jackie* (non-diagnostic), $M_{MPS-4} =$

415 5.60, $SD_{MPS-4} = 0.99$, *Charlie* (non-diagnostic), $M_{MPS-4} = 5.53$, $SD_{MPS-4} = 1.08$. For the
 416 diagnostic descriptions, there was significant variation depending on the description,
 417 $F(3,613) = 2.91$, $p = .036$, partial $\eta^2 = 0.00$, *Sam* was viewed significantly more favorably
 418 than *Francis* ($p = .040$). For the non-diagnostic descriptions there was no significant
 419 difference in ratings depending on description, $t(214) = -1.79$, $p = .075$, $d = 0.12$.

420 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
 421 (diagnostic), $M_{MM-1} = 79.85$, $SD_{MM-1} = 15.44$; *Francis* (diagnostic), $M_{MM-1} = 78.30$, SD_{MM-1}
 422 = 15.84; *Alex* (diagnostic), $M_{MM-1} = 79.78$, $SD_{MM-1} = 15.71$; *Robin* (diagnostic), $M_{MM-1} =$
 423 79.46, $SD_{MM-1} = 15.41$; *Jackie* (non-diagnostic), $M_{MM-1} = 73.44$, $SD_{MM-1} = 15.83$; *Charlie*
 424 (non-diagnostic), $M_{MM-1} = 73.07$, $SD_{MM-1} = 16.22$. For the diagnostic descriptions, we
 425 observed no significant variation depending on the description, $F(3,594) = 1.45$, $p = .231$,
 426 partial $\eta^2 = 0.002$. For the non-diagnostic descriptions there was no significant difference in
 427 ratings depending on description, $t(214) = -0.60$, $p = .552$, $d = 0.04$.

428 We conducted a linear-mixed-effects model to test if condition influenced MPS-4
 429 responses. Our outcome measure was MPS-4, our predictor variable was condition; we
 430 allowed intercepts and the effect of condition to vary across participants. Overall, the model
 431 significantly predicted participants responses, and provided a better fit for the data than the
 432 baseline model, $\chi^2(2) = 475.42$, $p < .001$. Condition was a significant predictor in the model
 433 $b = 0.19$, $t(214.35) = 6.53$, $p < .001$, with the diagnostic descriptions being rated as more
 434 moral than the non-diagnostic descriptions of immoral characters Figure 9.

435 We conducted a linear-mixed-effects model to test if condition influenced MM-1
 436 responses. Our outcome measure was MM-1, our predictor variable was condition; we
 437 allowed intercepts and the effect of condition to vary across participants. Overall, the model
 438 significantly predicted participants responses, and provided a better fit for the data than the
 439 baseline model, $\chi^2(2) = 324.13$, $p < .001$. Condition was a significant predictor in the model
 440 $b = 3.04$, $t(214.90) = 6.02$, $p < .001$, with the diagnostic descriptions being rated as more

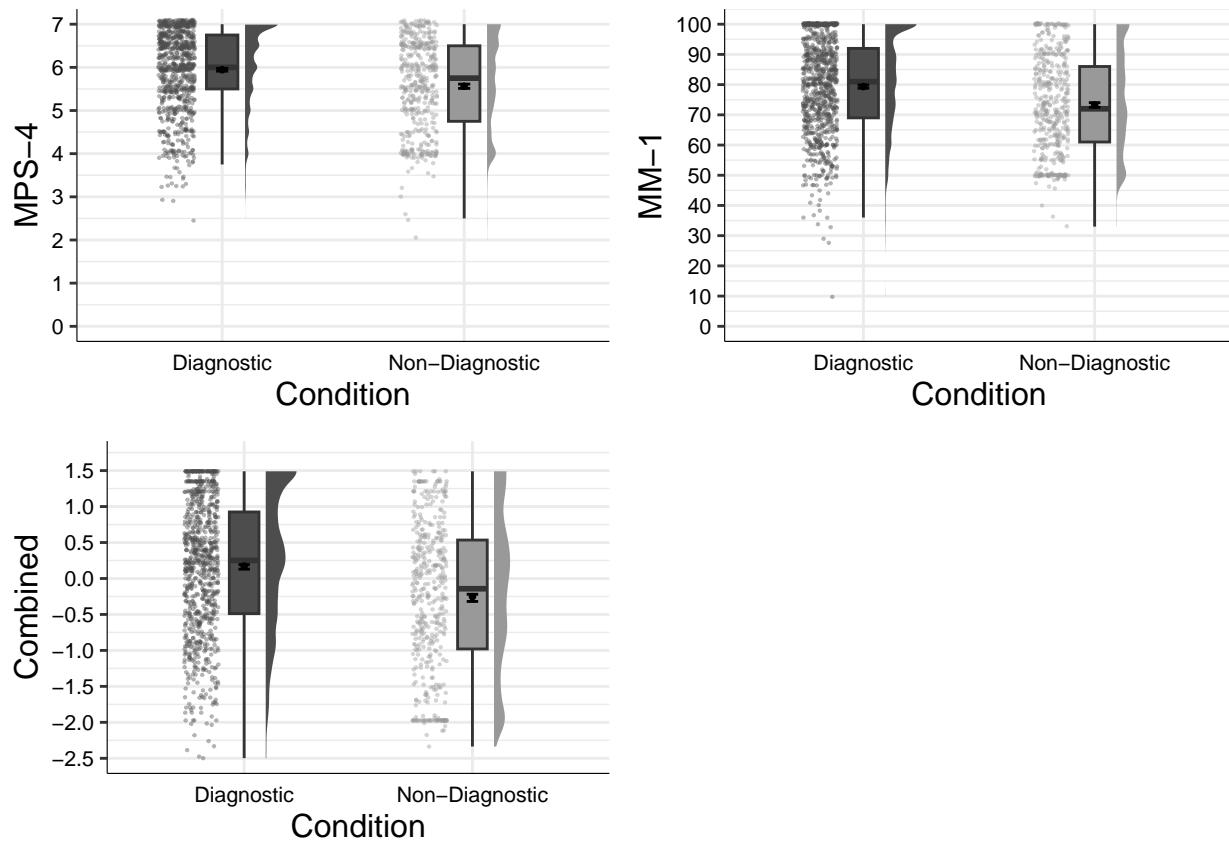


Figure 9. Pilot Study 2: Differences in moral perception depending on condition

⁴⁴¹ moral than the non-diagnostic descriptions, see Figure 9.

⁴⁴² **Pilot 2: Combined Measure.** As in previous studies, we developed a combined
⁴⁴³ moral perception measure by calculating the mean of the combined mean-centered scores for
⁴⁴⁴ MPS-4 and MM-1, and mean-centering this result. Below we report the analyses for this
⁴⁴⁵ combined measure.

⁴⁴⁶ The standardized means and standard deviations for the combined measure for each
⁴⁴⁷ scenario are as follows: *Sam* (moral), $M = 0.21$, $SD = 0.91$; *Francis* (moral), $M = 0.10$, SD
⁴⁴⁸ = 0.96; *Alex* (moral), $M = 0.18$, $SD = 0.94$; *Robin* (moral), $M = 0.16$, $SD = 0.93$; *Jackie*
⁴⁴⁹ (neutral), $M = -0.24$, $SD = 1.01$; *Charlie* (neutral), $M = -0.30$, $SD = 1.07$. For the moral
⁴⁵⁰ descriptions, we observed significant variation depending on the description, $F(3,588) = 2.90$,
⁴⁵¹ $p = .039$, 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).

459 Study S1 - Good Characters

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

461 **Study S1: Method**

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

468 A total sample of 1118 (642 female, 445 male, 2 non-binary, 3 other; 1 prefer not to say,
469 $M_{age} = 37.44$, min = 19, max = 84, $SD = 11.08$) started the survey. Participants were
470 recruited from MTurk and paid \$0.40 for their participation.

471 Participants who failed both manipulation checks were removed ($n = 262$), leaving a
472 total sample of 856 participants (507 female, 347 male, 0 other, 0 prefer not to say; $M_{age} =$
473 37.12, min = 19, max = 84, $SD = 11.04$).

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

476 **Study S1: Results**

477 **Study S1: Main Measures.** The means and standard deviations for MPS-4 for
478 each scenario are as follows: *Sam*, $M_{MPS-4} = 5.95$, $SD_{MPS-4} = 0.93$, *Francis*, $M_{MPS-4} = 5.89$,
479 $SD_{MPS-4} = 0.91$, *Alex*, $M_{MPS-4} = 5.94$, $SD_{MPS-4} = 0.96$, *Robin*, $M_{MPS-4} = 5.95$, $SD_{MPS-4} =$
480 0.94. There was significant variation depending on the description, $F(3,2527) = 3.30$, $p =$
481 .020, partial $\eta^2 = 0.001$. Pairwise comparisons did not reveal any significant differences
482 between individual descriptions (all $ps > .05$).

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

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

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

Study S1: Combined Measure. The means and standard deviations for the combined measure for each scenario are as follows: *Sam*, $M = 0.03$, $SD = 1.02$, *Francis*, $M = -0.03$, $SD = 0.98$, *Alex*, $M = 0.02$, $SD = 1.04$, *Robin*, $M = 0.04$, $SD = 1.01$. There was significant variation depending on the description, $F(3,2493) = 4.32$, $p = .005$, partial $\eta^2 =$

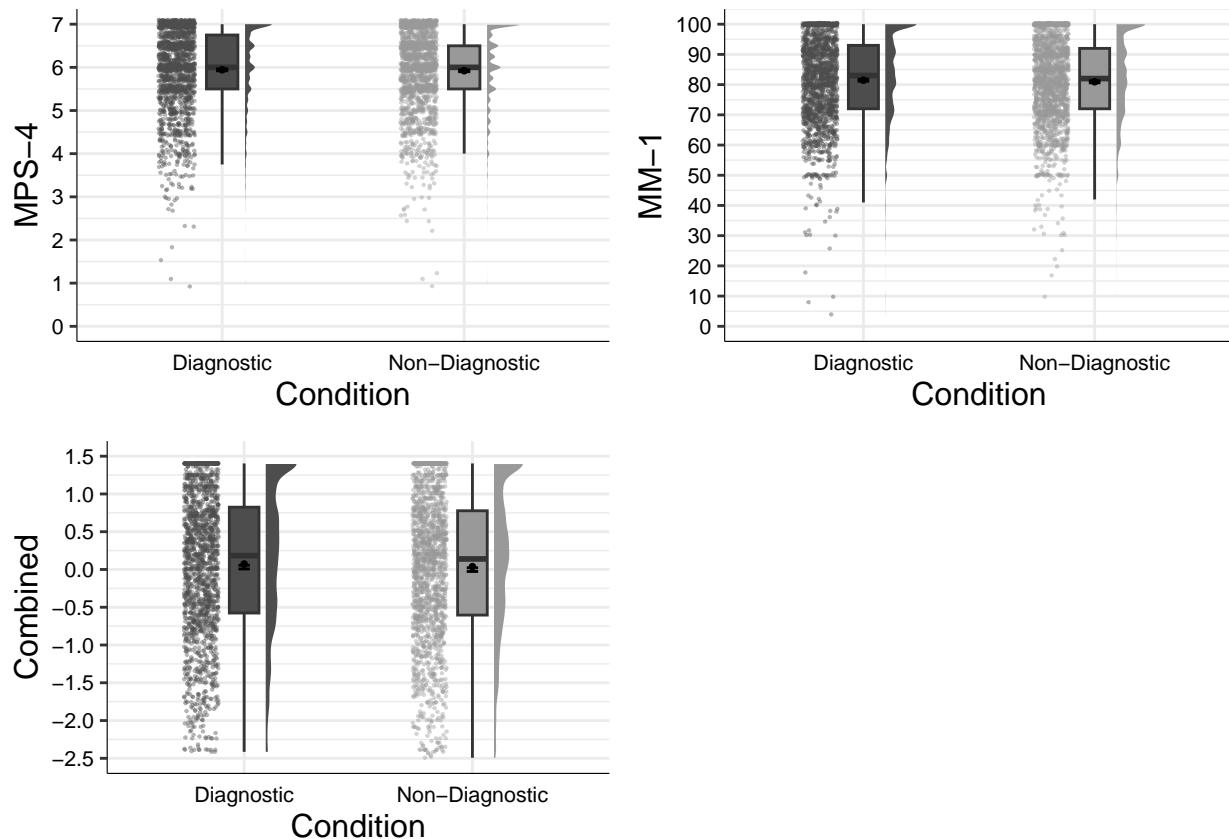


Figure 10. Study S1: Responses to moral perception measures depending on condition

509 0.00. Follow-up pairwise comparisons did not reveal any significant differences between the
 510 different characters (all $p > .05$).

511 We conducted a linear-mixed-effects model to test if condition influenced moral
 512 perception. Our outcome measure was the combined moral perception measure, our
 513 predictor variable was condition; we allowed intercepts and the effect of condition to vary
 514 across participants, and scenario was also included in the model. Overall, the model
 515 significantly predicted participants responses, and provided a better fit for the data than the
 516 baseline model, $\chi^2(8) = 42.42$, $p < .001$. Condition did not influence moral perception, $F(1,$
 517 $865.01) = 5.31$, $p = .021$; and was not a significant predictor in the model when controlling
 518 for scenario, $b = -0.01$, $t(2,541.03) = -0.82$, $p = .410$, see Figure 4.

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

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

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

523 ($M = 5.89, SD = 0.91$), than in the diagnostic condition ($M = 6.02, SD = 0.95$), $t(810.53) =$
524 $1.97, p = .049, d = 0.14$; MM-1 ratings were similar in the non-diagnostic condition ($M =$
525 $79.75, SD = 14.62$), than in the diagnostic condition ($M = 83.25, SD = 13.30$), $t(845.88) =$
526 $3.66, p < .001, d = 0.25$. For the combined measure ratings were also similar in the
527 non-diagnostic condition ($M = -0.06, SD = 1.03$), than in the diagnostic condition ($M =$
528 $0.15, SD = 1.01$), $t(829.20) = 3.07, p = .002, d = 0.21$.

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

530 condition ($M = 5.95, SD = 0.93$), than in the diagnostic condition ($M = 5.94, SD = 0.95$),
531 $t(811.83) = -0.20, p = .841, d = 0.01$; MM-1 ratings were similar in the non-diagnostic
532 condition ($M = 81.62, SD = 14.28$), and in the diagnostic condition ($M = 81.64, SD =$
533 14.02), $t(824.54) = 0.02, p = .982, d = 0.00$. For the combined measure ratings were also
534 similar in the non-diagnostic condition ($M = 0.04, SD = 1.03$), than in the diagnostic
535 condition ($M = 0.04, SD = 0.99$), $t(828.47) = -0.10, p = .919, d = 0.01$.

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

537 ($M = 5.97, SD = 0.91$), than in the diagnostic condition ($M = 5.91, SD = 0.99$), $t(845.29) =$
538 $-0.91, p = .362, d = 0.06$; MM-1 ratings were similar in the non-diagnostic condition ($M =$
539 $81.93, SD = 13.38$), than in the diagnostic condition ($M = 80.51, SD = 15.21$), $t(850.53) =$
540 $-1.46, p = .145, d = 0.10$. For the combined measure ratings were also similar in the
541 non-diagnostic condition ($M = 0.07, SD = 0.98$), than in the diagnostic condition ($M =$
542 $-0.02, SD = 1.09$), $t(847.27) = -1.30, p = .192, d = 0.09$.

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

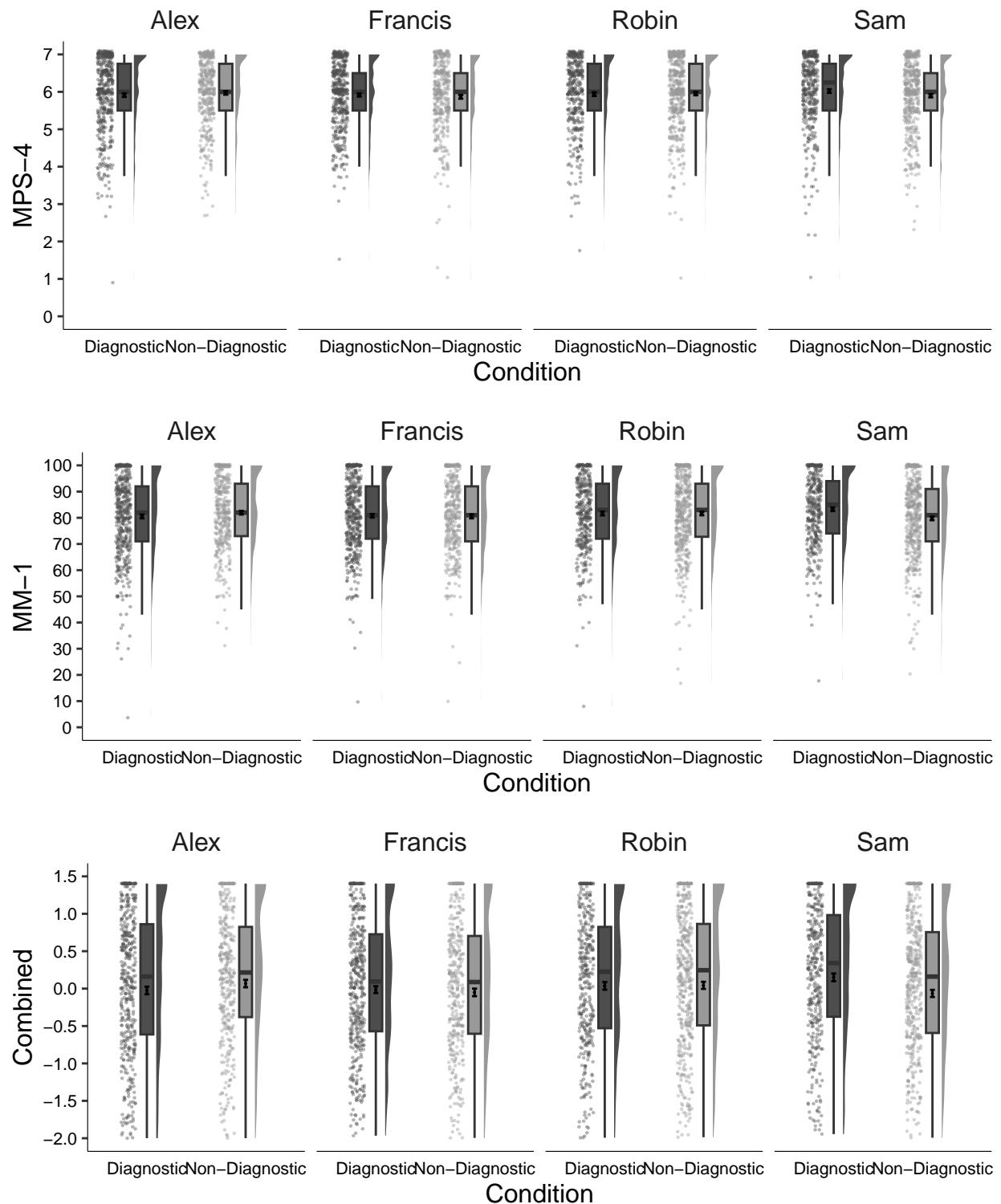


Figure 11. Study 2: Differences in moral perception for each description

544 condition ($M = 5.87, SD = 0.95$), than in the diagnostic condition ($M = 5.91, SD = 0.87$),
545 $t(787.36) = 0.77, p = .443, d = 0.05$; MM-1 ratings were not significantly different in the
546 non-diagnostic condition ($M = 80.54, SD = 14.38$), than in the diagnostic condition ($M =$
547 $80.75, SD = 13.99$), $t(809.63) = 0.21, p = .832, d = 0.01$. For the combined measure ratings
548 were also similar in the non-diagnostic condition ($M = -0.05, SD = 0.99$), and in the
549 diagnostic condition ($M = -0.01, SD = 0.98$), $t(814.30) = 0.55, p = .581, d = 0.04$.

Study S2 - Good and Bad Characters

550
551 Study S2 is the same as Study 3, but with an MTurk sample. Study 3 was
552 pre-registered at https://aspredicted.org/QDF_XT1.

553 **Study S2: Method**

554 **Study S2: Participants and design.** Study S2 was a 2×2 within-subjects
555 factorial design. The first independent variable was condition with two levels, diagnostic and
556 non-diagnostic. The second independent variable was valence of character description, with
557 two levels morally good and morally bad. We used the same two dependent variables as in
558 previous studies, the four item moral perception scale (MPS-4, $\alpha = 0.94$), and the single
559 item moral perception measure MM-1.

560 A total sample of 1095 (700 female, 386 male, 2 non-binary, 0 other; 2 prefer not to say,
561 $M_{age} = 36.42$, min = 19, max = 77, $SD = 10.65$) started the survey. Participants were
562 recruited from MTurk and paid \$0.40 for their participation.

563 Participants who failed both manipulation checks were removed ($n = 221$), leaving a
564 total sample of 874 participants (550 female, 320 male, 0 other, 0 prefer not to say; $M_{age} =$
565 36.37, min = 19, max = 77, $SD = 10.72$).

566 **Study S2: Procedure and materials.** Again, data were collected using an online
567 questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented
568 with four descriptions of characters as in Study 3. To ensure consistency across character
569 judgments, we selected descriptions that related to the same moral foundations (care,
570 fairness, and loyalty). We used the same four character names as in previous studies. The
571 *good* characters were *Sam* and *Robin*, and the *bad* characters were *Francis* and *Alex*, e.g.,
572 *Imagine a person named Robin. Throughout their life they have been known to show*
573 *compassion and empathy for others, act with a sense of fairness and justice, and, never to*
574 *break their word.* or, *Imagine a person named Alex. Throughout their life they have been*

575 known to be cruel, act unfairly, and to betray their own group. Full descriptions for each
576 character are in the supplementary materials. One description for each the *good* and *bad*
577 characters was randomly assigned to include non-diagnostic information for each participant
578 thus all participants were exposed to all conditions (see
579 https://osf.io/mdnpv/?view_only=77883e3fb3d45f1a35fe92d5318cb67 for details of the
580 randomization blocks). Study S2 was pre-registered at https://aspredicted.org/QDF_XT1

581 **Study S2: Results**

582 The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*
583 (*good*), $M_{MPS-4} = 5.90$, $SD_{MPS-4} = 1.03$, *Francis* (*bad*), $M_{MPS-4} = 4.07$, $SD_{MPS-4} = 2.07$, *Alex*
584 (*bad*), $M_{MPS-4} = 4.03$, $SD_{MPS-4} = 2.03$, *Robin* (*good*), $M_{MPS-4} = 5.85$, $SD_{MPS-4} = 1.05$. There
585 was significant variation depending on the description, $F(1,1080) = 442.71$, $p < .001$, partial
586 $\eta^2 = 0.24$. Both the *good* characters (*Robin* and *Sam*) were rated significantly more favorably
587 than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$). There were no differences
588 between *Robin* and *Sam* (*good*: $p = .366$) or between *Alex* and *Francis* (*bad*; $p = .648$).

589 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
590 (*good*), $M_{MM-1} = 81.01$, $SD_{MM-1} = 15.23$; *Francis* (*bad*), $M_{MM-1} = 51.49$, $SD_{MM-1} = 33.18$;
591 *Alex* (*bad*), $M_{MM-1} = 50.89$, $SD_{MM-1} = 32.14$; *Robin* (*good*), $M_{MM-1} = 80.81$, $SD_{MM-1} =$
592 15.16. There was significant variation depending on the description, $F(1,1080) = 458.92$, $p <$
593 .001, partial $\eta^2 = 0.254$. Again, the *good* characters (*Robin* and *Sam*) were rated
594 significantly more favorably than the *bad* characters (*Alex* and *Francis*; all $ps < .001$). There
595 were no differences between *Robin* and *Sam* (*good*: $p = .776$) or between *Alex* and *Francis*
596 (*bad*; $p = .683$).

597 We conducted a linear-mixed-effects model to test if our predictors influenced MPS-4
598 responses. Our outcome measure was MPS-4, our predictor variables were condition and
599 valence; we allowed intercepts and the effects of condition and valence to vary across

600 participants. Overall, the model significantly predicted participants responses, and provided
601 a better fit for the data than the baseline model, $\chi^2(5) = 4,554.31, p < .001$. Overall, there
602 was a significant main effect for condition, $F(1, 873) = 8.61, p = .003$; valence significantly
603 predicted responses, $F(1, 873) = 1,859.34, p < .001$; and there was no significant condition
604 \times valence interaction, $F(1, 873) = 0.01, p = .935$.

605 We conducted a linear-mixed-effects model to test if our predictors influenced MM-1
606 responses. The model was the same as the previous model, with a change to the outcome
607 measure, our outcome measure for this model was MM-1. As above, our predictor variables
608 were condition and valence; we allowed intercepts and the effects of condition and valence to
609 vary across participants. Overall, the model significantly predicted participants responses,
610 and provided a better fit for the data than the baseline model, $\chi^2(5) = 3,496.86, p < .001$.
611 Overall there was a main effect for condition, $F(1, 873) = 16.61, p < .001$; valence
612 significantly predicted responses, $F(1, 873) = 986.35, p < .001$; and there was no significant
613 condition \times valence interaction, $F(1, 873) = 0.04, p = .849$.

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

624 For both MP-4 and MM-1 (and the combined measure) we found a main effect for
625 condition and valence, and there was no condition \times valence interaction. We conducted

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

628 **Differences in the *Bad* Descriptions**

629 For the *bad* characters, we conducted a linear-mixed-effects model to test if condition
630 influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was
631 condition; we allowed intercepts and the effect of condition to vary across participants.
632 Overall, the model did not significantly predict participants responses, or provide a better fit
633 for the data than the baseline model, $\chi^2(3) = 5.40, p = .145$. Condition did not significantly
634 influence MPS-4 responses $F(1, 872.00) = 3.54, p = .060$, and was not a significant predictor
635 in the model $b = -0.03, t(872.00) = -1.88, p = .060$, see Figure 12.

636 We also conducted a linear-mixed-effects model to test if condition influenced MM-1
637 responses. Our outcome measure was MM-1, our predictor variable was condition; we
638 allowed intercepts and the effect of condition to vary across participants. Overall, the model
639 significantly predicted participants responses, and provided a better fit for the data than the
640 baseline model, $\chi^2(3) = 8.67, p = .034$. Condition significantly influenced MM-1 responses
641 $F(1, 872.00) = 7.01, p = .008$, and was a significant predictor in the model $b = -0.69,$
642 $t(872.00) = -2.65, p = .008$, see Figure 12.

643 **Differences in the *Good* Descriptions**

644 For the *good* characters, we conducted a linear-mixed-effects model to test if condition
645 influenced MPS-4 responses. Our outcome measure was MPS-4, our predictor variable was
646 condition; we allowed intercepts and the effect of condition to vary across participants.
647 Overall, the model significantly predicted participants responses, and provided a better fit
648 for the data than the baseline model, $\chi^2(3) = 13.66, p = .003$. Condition significantly
649 influenced MPS-4 responses $F(1, 872.00) = 6.82, p = .009$, and was a significant predictor in
650 the model $b = 0.03, t(872.00) = 2.61, p = .009$, see Figure 12.

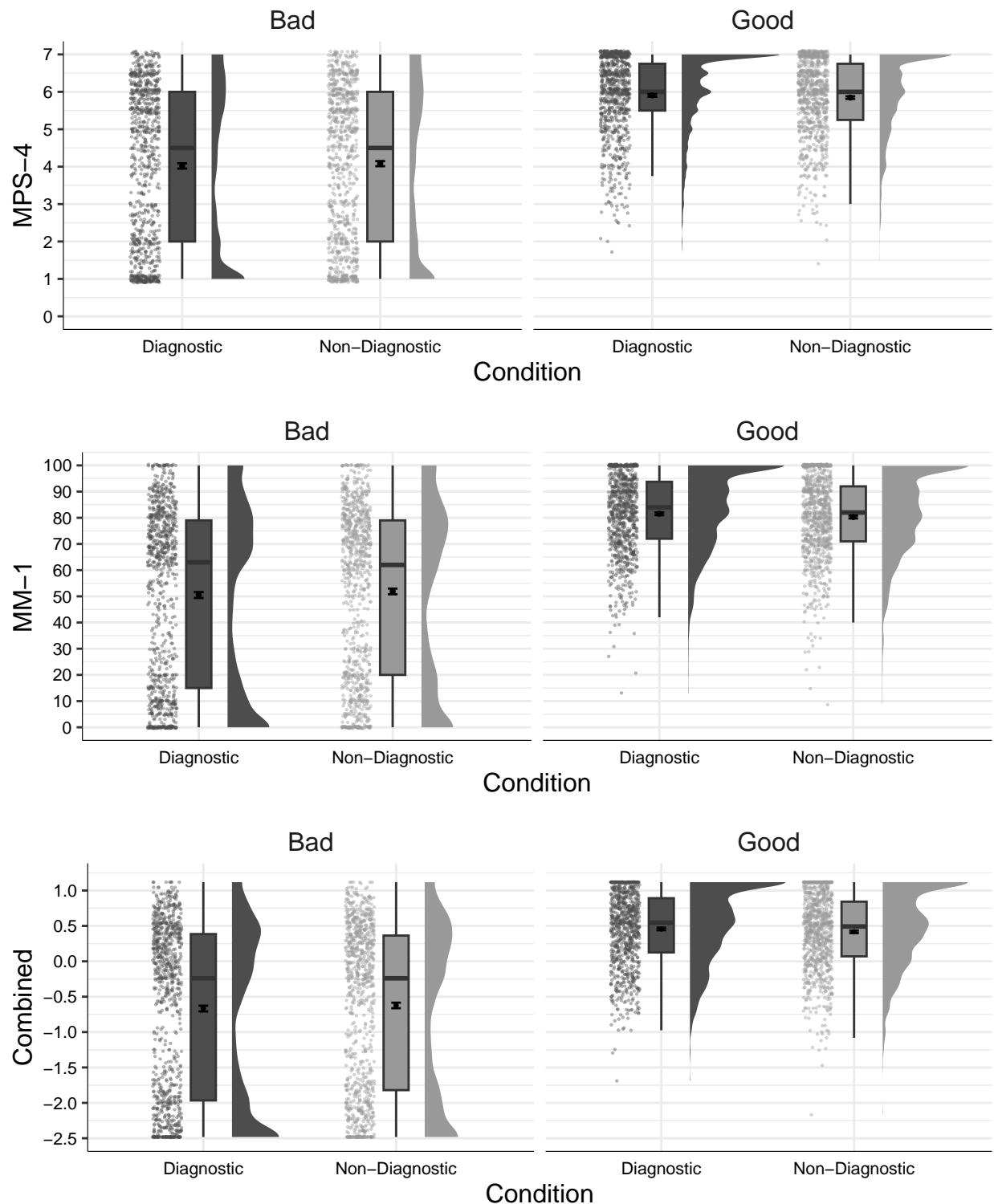


Figure 12. Study S2: Differences in moral perception depending on condition

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, $\chi^2(1) = 11.97, p < .001$. Condition significantly influenced MM-1 responses $F(1, 873) = 12.04, p < .001$, and was a significant predictor in the model $b = 0.63, t(873) = 3.47, p < .001$, see Figure 12.

Study S2: Differences between the descriptions

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

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

For *Robin (good)*, MPS-4 scores were not significantly different for the non-diagnostic condition ($M = 5.88, SD = 0.96$), than in the diagnostic condition ($M = 5.83, SD = 1.14$), $t(844.53) = -0.77, p = .440, d = 0.05$; MM-1 ratings were similar in the non-diagnostic condition ($M = 80.92, SD = 15.27$), and in the diagnostic condition ($M = 80.70, SD = 15.07$), $t(871.98) = -0.22, p = .828, d = 0.01$. For the combined measure ratings were also similar in the non-diagnostic condition ($M = 0.44, SD = 0.51$), than in the diagnostic condition ($M = 0.42, SD = 0.54$), $t(867.63) = -0.57, p = .569, d = 0.04$.

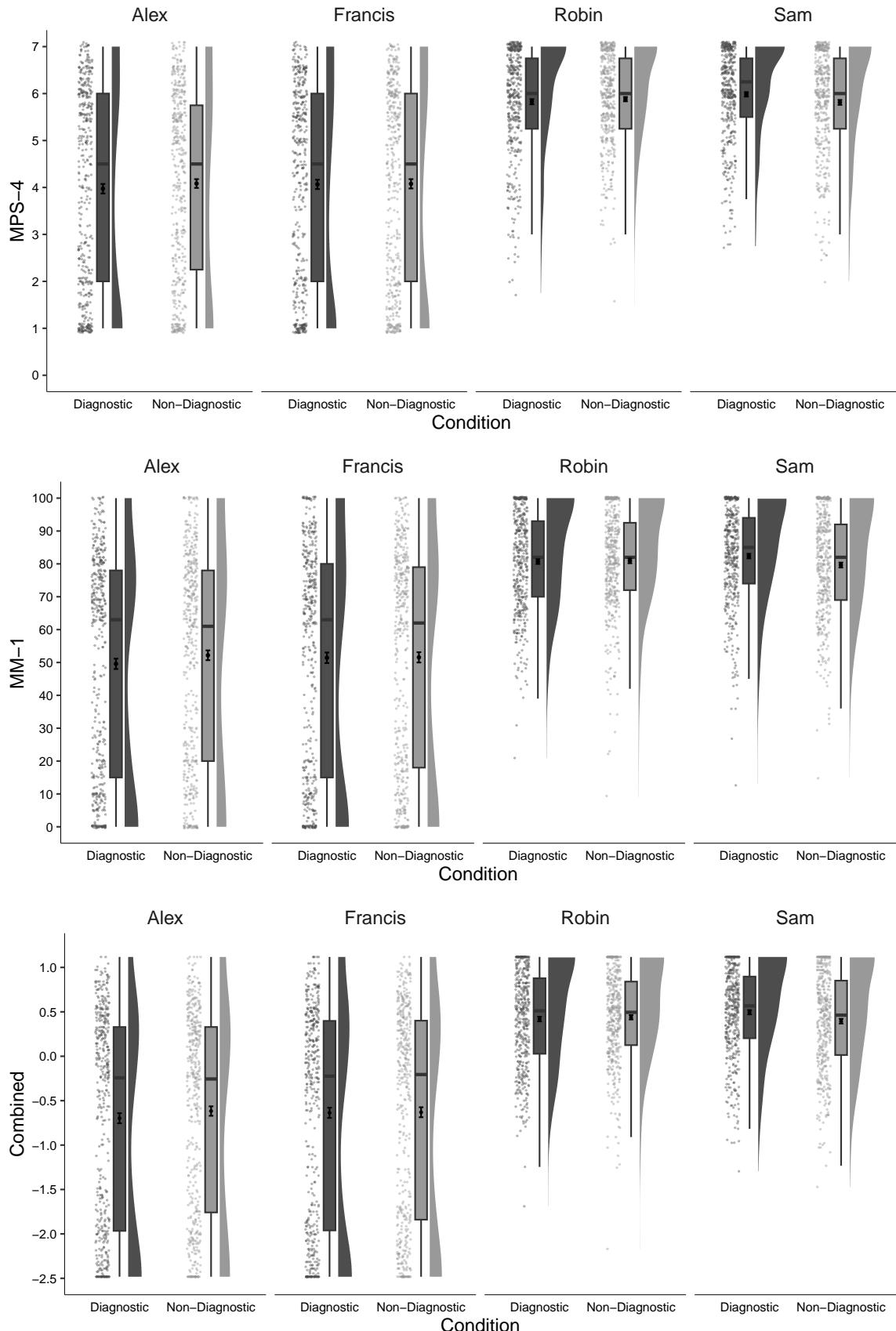


Figure 13. Study 3: Differences in moral perception for each description

676 For *Alex (bad)*, MPS-4 scores were not significantly different for the non-diagnostic
677 condition ($M = 4.08$, $SD = 1.96$), than in the diagnostic condition ($M = 3.97$, $SD = 2.11$),
678 $t(865.81) = -0.80$, $p = .421$, $d = 0.05$; MM-1 ratings were similar in the non-diagnostic
679 condition ($M = 52.19$, $SD = 31.29$), and in the diagnostic condition ($M = 49.58$, $SD =$
680 32.95), $t(868.76) = -1.20$, $p = .230$, $d = 0.08$. For the combined measure ratings were also
681 similar in the non-diagnostic condition ($M = -0.62$, $SD = 1.11$), and in the diagnostic
682 condition ($M = -0.70$, $SD = 1.19$), $t(867.67) = -1.04$, $p = .301$, $d = 0.07$.

683 For *Francis (bad)*, MPS-4 scores were not significantly different for the non-diagnostic
684 condition ($M = 4.08$, $SD = 2.07$), than in the diagnostic condition ($M = 4.07$, $SD = 2.07$),
685 $t(871.94) = -0.09$, $p = .928$, $d = 0.01$; MM-1 ratings were not significantly different in the
686 non-diagnostic condition ($M = 51.56$, $SD = 32.68$), than in the diagnostic condition ($M =$
687 51.42 , $SD = 33.70$), $t(871.59) = -0.06$, $p = .952$, $d = 0.00$. For the combined measure ratings
688 were also similar in the non-diagnostic condition ($M = -0.63$, $SD = 1.18$), and in the
689 diagnostic condition ($M = -0.64$, $SD = 1.20$), $t(871.88) = -0.08$, $p = .939$, $d = 0.01$.

690

Study S3 - Good and Bad Characters

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

694 **Study S3: Method**

695 **Study S3: Participants and design.** Study S3 was a 2×2 between-subjects
696 factorial design. As in Study 3, the first independent variable was condition with two levels,
697 diagnostic and non-diagnostic. The second independent variable was valence of character
698 description, with two levels morally good and morally bad. We used the same two dependent
699 variables as in previous studies (MPS-4, $\alpha = 0.97$, and MM-1).

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

703 Participants who failed both manipulation checks were removed ($n = 445$), leaving a
704 total sample of 1944 participants (970 female, 960 male, 2 other, 2 prefer not to say; $M_{age} =$
705 37.88, min = 2, max = 454, $SD = 15.49$).

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

713 **Study S3: Results**

714 The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*
715 (*good*), $M_{MPS-4} = 6.15$, $SD_{MPS-4} = 0.87$, *Francis* (*bad*), $M_{MPS-4} = 3.65$, $SD_{MPS-4} = 2.16$, *Alex*
716 (*bad*), $M_{MPS-4} = 3.65$, $SD_{MPS-4} = 2.09$, *Robin* (*good*), $M_{MPS-4} = 6.21$, $SD_{MPS-4} = 0.85$. There
717 was significant variation depending on the description, $F(3,1940) = 396.86$, $p < .001$, partial
718 $\eta^2 = 0.38$. Both the *good* characters (*Robin* and *Sam*) were rated significantly more favorably
719 than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$). There were no differences
720 between *Robin* and *Sam* (*good*: $p = .932$) or between *Alex* and *Francis* (*bad*; $p > .999$).

721 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
722 (*good*), $M_{MM-1} = 84.70$, $SD_{MM-1} = 15.32$; *Francis* (*bad*), $M_{MM-1} = 43.37$, $SD_{MM-1} = 34.96$;
723 *Alex* (*bad*), $M_{MM-1} = 44.68$, $SD_{MM-1} = 34.57$; *Robin* (*good*), $M_{MM-1} = 85.33$, $SD_{MM-1} =$
724 14.47. There was significant variation depending on the description, $F(3,1940) = 383.99$, $p <$
725 $.001$, partial $\eta^2 = 0.37$. Both the *good* characters (*Robin* and *Sam*) were rated significantly
726 more favorably than both the *bad* characters (*Alex* and *Francis*; all $ps < .001$). There were
727 no differences between *Robin* and *Sam* (*good*: $p = .982$) or between *Alex* and *Francis* (*bad*;
728 ($p = .872$)).

729 We conducted a 2×2 between subjects ANOVA to test for an interaction between
730 valence and condition in predicting MPS-4. Condition significantly influenced responses to
731 the MPS-4, $F(1, 1940) = 5.16$, $p = .023$; valence significantly predicted responses, $F(1, 1940)$
732 = 1,495.09, $p < .001$; and there was no significant condition \times valence interaction, $F(1,$
733 1940) = 0.03, $p = .858$.

734 We conducted a 2×2 between subjects ANOVA to test for an interaction between
735 valence and condition in predicting responses to MM-1. Condition significantly influenced
736 responses to MM-1, $F(1, 1940) = 9.46$, $p = .002$; valence significantly predicted responses,
737 $F(1, 1940) = 580.03$, $p < .001$; and there was no significant condition \times valence interaction,

738 $F(1, 1940) = 0.32, p = .573.$

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

744 As in previous studied we conducted separate analyses for the good and bad
745 descriptions.

746 Differences in the *Bad* Descriptions

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

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

753 Differences in the *Good* Descriptions

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

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

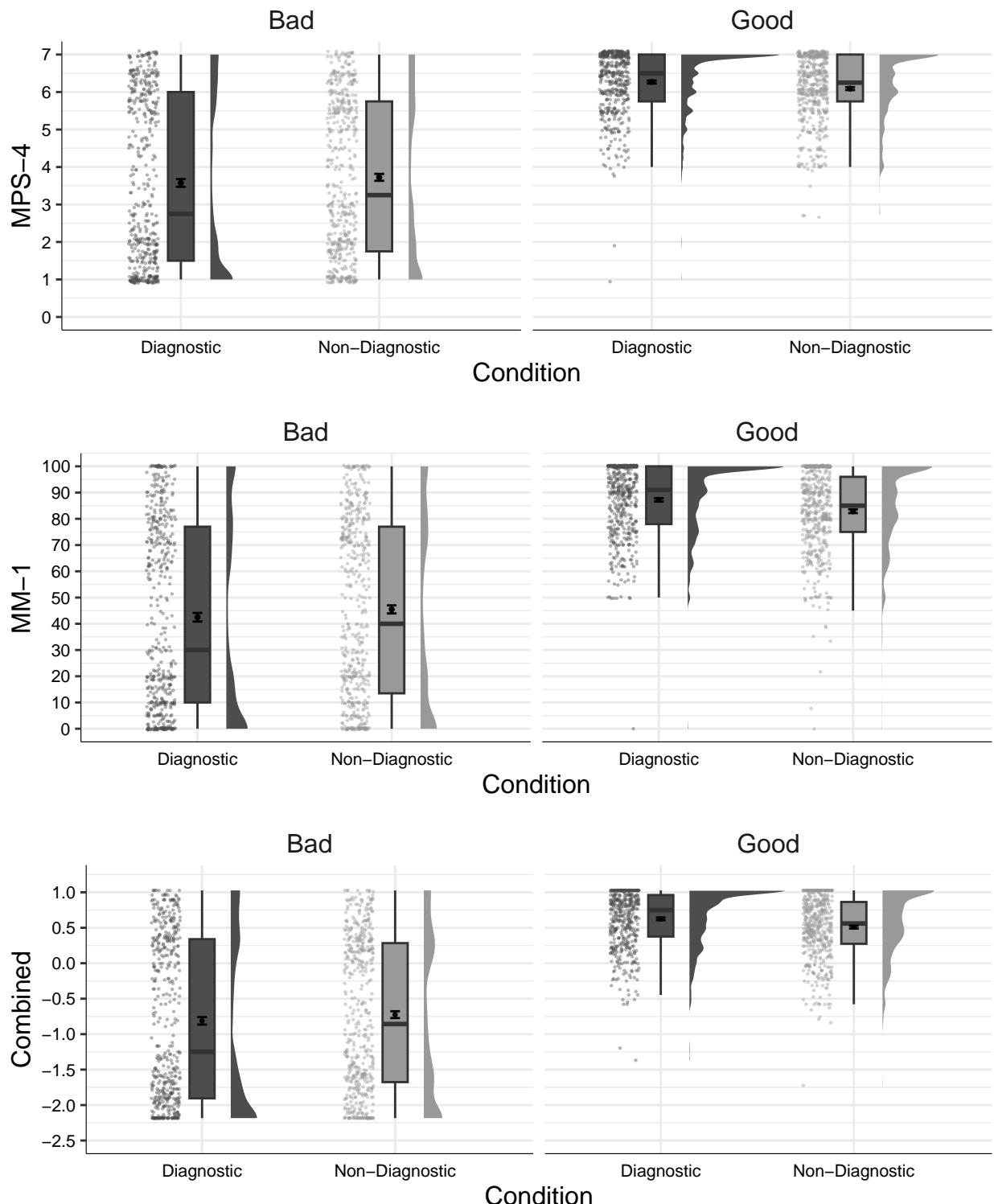


Figure 14. Study S3: Differences in moral perception depending on condition

760 **Study S3: Differences between the descriptions**

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

764 For *Sam*, MPS-4 scores were significantly lower in the non-diagnostic condition ($M =$
765 $6.06, SD = 0.89$), than in the diagnostic condition ($M = 6.24, SD = 0.85$), $t(488.52) = 2.34$,
766 $p = .020, d = 0.21$; Similarly, MM-1 ratings were significantly lower in the non-diagnostic
767 condition ($M = 82.23, SD = 16.63$), than in the diagnostic condition ($M = 87.21, SD =$
768 13.43), $t(471.87) = 3.65, p < .001, d = 0.33$. For the combined measure ratings were also
769 lower in the non-diagnostic condition ($M = 0.49, SD = 0.45$), than in the diagnostic
770 condition ($M = 0.62, SD = 0.41$), $t(486.39) = 3.35, p < .001, d = 0.30$.

771 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
772 condition ($M = 6.13, SD = 0.87$), than in the diagnostic condition ($M = 6.30, SD = 0.83$),
773 $t(490.93) = 2.21, p = .027, d = 0.20$; MM-1 ratings were similar in the non-diagnostic
774 condition ($M = 83.53, SD = 14.75$), and in the diagnostic condition ($M = 87.25, SD =$
775 13.96), $t(490.98) = 2.88, p = .004, d = 0.26$. For the combined measure ratings were also
776 similar in the non-diagnostic condition ($M = 0.53, SD = 0.42$), than in the diagnostic
777 condition ($M = 0.63, SD = 0.40$), $t(490.92) = 2.82, p = .005, d = 0.25$.

778 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic condition
779 ($M = 3.71, SD = 2.02$), than in the diagnostic condition ($M = 3.59, SD = 2.16$), $t(465.98) =$
780 $-0.62, p = .534, d = 0.06$; MM-1 ratings were similar in the non-diagnostic condition ($M =$
781 $45.80, SD = 33.76$), than in the diagnostic condition ($M = 43.48, SD = 35.46$), $t(468.21) =$
782 $-0.73, p = .465, d = 0.07$. For the combined measure ratings were also similar in the
783 non-diagnostic condition ($M = -0.72, SD = 1.05$), than in the diagnostic condition ($M =$
784 $-0.79, SD = 1.13$), $t(465.61) = -0.69, p = .489, d = 0.06$.

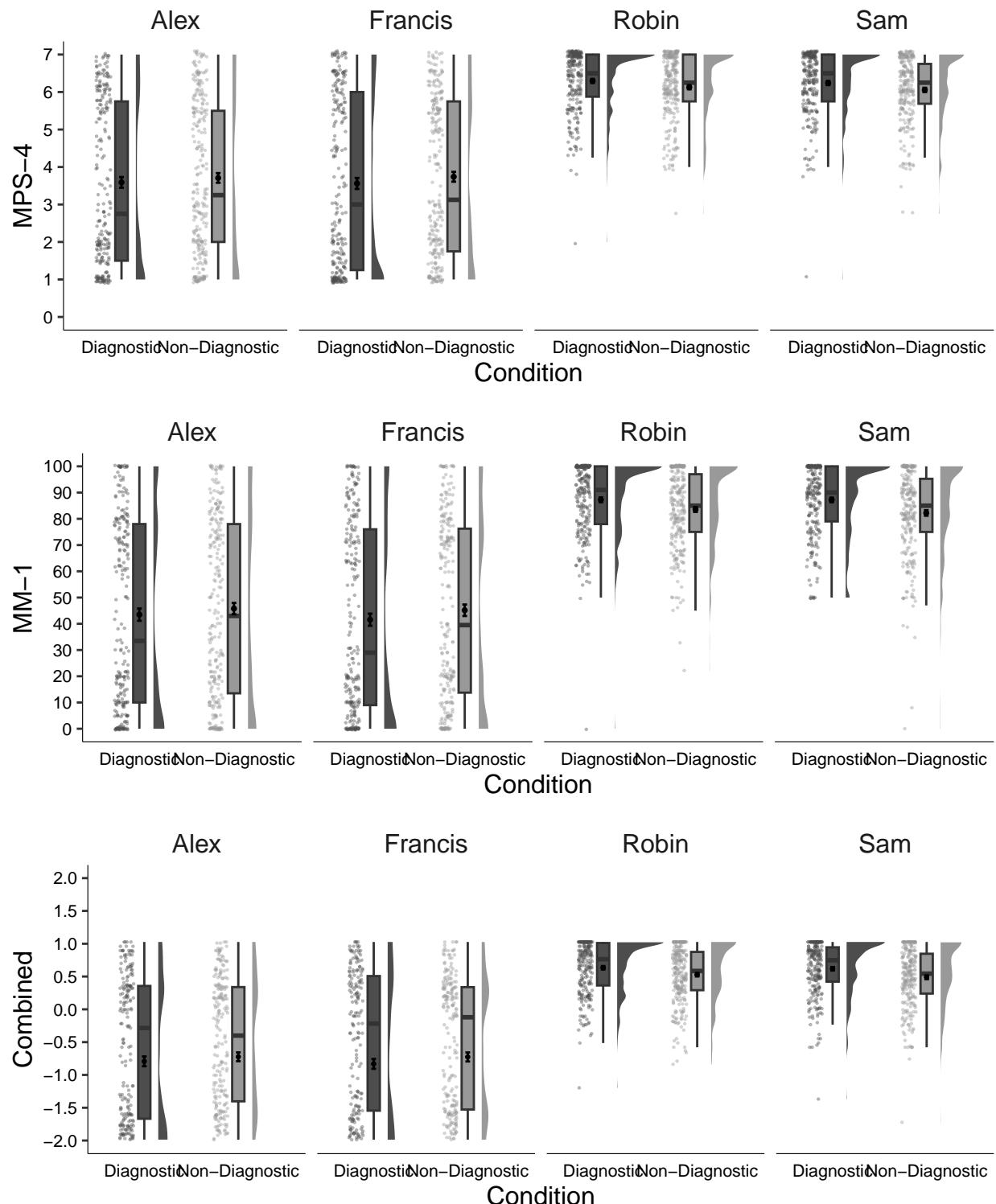


Figure 15. Study S3: Differences in moral perception for each description

785 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
786 condition ($M = 3.74$, $SD = 2.05$), than in the diagnostic condition ($M = 3.56$, $SD = 2.27$),
787 $t(474.15) = -0.91$, $p = .364$, $d = 0.08$; MM-1 ratings were not significantly different in the
788 non-diagnostic condition ($M = 45.16$, $SD = 34.18$), than in the diagnostic condition ($M =$
789 41.55 , $SD = 35.73$), $t(478.97) = -1.13$, $p = .258$, $d = 0.10$. For the combined measure ratings
790 were also similar in the non-diagnostic condition ($M = -0.73$, $SD = 1.07$), and in the
791 diagnostic condition ($M = -0.83$, $SD = 1.16$), $t(476.42) = -1.04$, $p = .297$, $d = 0.10$.

792

References

- 793 Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., ... Eden,
794 A. (2020). Validating the extended character morality questionnaire. *Media Psychology*,
795 23(1), 107–130. <https://doi.org/10.1080/15213269.2019.1572523>
- 796 Walker, A. C., Turpin, M. H., Fugelsang, J. A., & Białek, M. (2021). Better the two devils
797 you know, than the one you don't: Predictability influences moral judgments of immoral
798 actors. *Journal of Experimental Social Psychology*, 97, 104220.
799 <https://doi.org/10.1016/j.jesp.2021.104220>