

¹ Supplement: Moral Dilution² Cillian McHugh¹ & Eric R. Igou¹³ ¹ University of Limerick⁴ Author Note⁵ Department of Psychology, University of Limerick.

⁶ All procedures performed in studies involving human participants were approved by
⁷ institutional research ethics committee and conducted in accordance with the Code of
⁸ Professional Ethics of the Psychological Society of Ireland, and with the 1964 Helsinki
⁹ declaration and its later amendments or comparable ethical standards. Informed consent
¹⁰ was obtained from all individual participants included in the study. The authors declare
¹¹ that there are no potential conflicts of interest with respect to the research, authorship,
¹² and/or publication of this article. All authors consented to the submission of this
¹³ manuscript.

¹⁴ Correspondence concerning this article should be addressed to Cillian McHugh,
¹⁵ University of Limerick, Limerick, Ireland, V94 T9PX. E-mail: cillian.mchugh@ul.ie

16

Abstract

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

19 *Keywords:* keywords

20 Word count: X

21 Supplement: Moral Dilution

22 **Pilot Study 1**

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

27 **Pilot Study 1: Method**

28 **Pilot 1: Participants and design.** The pilot study was a within-subjects design.
29 The independent variable was description type with two levels, *diagnostic* and
30 *non-diagnostic*. We used two dependent variables. The first dependent variable was the
31 four item moral perception scale (MPS-4), participants rated the characters on four
32 dimensions using 7-point bipolar scales. The dimensions and scale endpoints were:
33 Bad-Good, Immoral-Moral, Violent-Peaceful, Merciless-Empathetic, this showed excellent
34 reliability, $\alpha = 0.93$. The second dependent variable was a single item moral perception
35 measure (MM-1) which consisted of a 100-point slider ranging from 0 = *Very Immoral* to
36 100 = *Very Moral*. Both dependent variables were taken from Walker et al. (2021).

37 A total sample of 235 (89 female, 142 male, 1 non-binary, 1 prefer not to say; $M_{age} =$
38 36.45, min = 20, max = 72, $SD = 10.23$) started the survey. Participants were recruited
39 from MTurk.

40 We removed participants who failed both manipulation checks ($n = 23$), leaving a
41 total sample of 212 participants (80 female, 128 male, 1 non-binary, 1 prefer not to say;
42 $M_{age} = 36.63$, min = 20, max = 72, $SD = 10.34$).

43 **Pilot 1: Procedure and materials.** Data were collected using an online
44 questionnaire presented with Qualtrics (www.qualtrics.com). Participants were presented

45 with descriptions of six characters.

46 Moral character descriptions were developed by combining descriptions relating to
47 three different moral foundations. These descriptions were adapted from the items of the
48 extended character morality questionnaire (Grizzard et al., 2020), and read as follows: (i)
49 *Imagine a person named Sam. Throughout their life they have been known to be cruel, act*
50 *unfairly, and to betray their own group;* (ii) *Imagine a person named Robin. Throughout*
51 *their life they have been known to physically hurt others, treat some people differently to*
52 *others, and show lack of loyalty;* (iii) *Imagine a person named Francis. Throughout their*
53 *life they have been known to violate the standards of purity and decency, show lack of*
54 *respect for authority, and treat people unequally* (iv) *Imagine a person named Alex.*
55 *Throughout their life they have been known to cause others to suffer emotionally, to deny*
56 *others their rights, and to cause chaos or disorder.*

57 We developed neutral descriptions that included information relating to physical
58 appearance/attributes, hobbies/activities, and family information that read as follows: (i)
59 *Imagine a person named Jackie. They have red hair, play tennis four times a month, and*
60 *have one older sibling and one younger sibling;* (ii) *Imagine a person named Charlie. They*
61 *are left-handed, drink tea in the morning, and have two older siblings and one younger*
62 *sibling.*

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

67 Pilot 1: Results

68 The means and standard deviations for MPS-4 for each scenario are as follows: *Sam*
69 (*diagnostic*), $M_{MPS-4} = 4.35$, $SD_{MPS-4} = 1.90$, *Francis* (*diagnostic*), $M_{MPS-4} = 4.46$,

70 $SD_{MPS-4} = 1.73$, *Alex* (diagnostic), $M_{MPS-4} = 4.44$, $SD_{MPS-4} = 1.79$, *Robin* (diagnostic),
 71 $M_{MPS-4} = 4.35$, $SD_{MPS-4} = 1.96$, *Jackie* (non-diagnostic), $M_{MPS-4} = 5.40$, $SD_{MPS-4} = 1.01$,
 72 *Charlie* (non-diagnostic), $M_{MPS-4} = 5.38$, $SD_{MPS-4} = 1.01$. For the diagnostic descriptions,
 73 there was no significant variation depending on the description, $F(3,600) = 1.58$, $p = .194$,
 74 partial $\eta^2 = 0.00$. For the non-diagnostic descriptions there was no significant difference in
 75 ratings depending on description, $t(211) = -0.67$, $p = .506$, $d = 0.05$.

76 The means and standard deviations for MM-1 for each scenario are as follows: *Sam*
 77 (diagnostic), $M_{MM-1} = 55.67$, $SD_{MM-1} = 30.47$; *Francis* (diagnostic), $M_{MM-1} = 58.22$,
 78 $SD_{MM-1} = 28.61$; *Alex* (diagnostic), $M_{MM-1} = 56.80$, $SD_{MM-1} = 29.45$; *Robin* (diagnostic),
 79 $M_{MM-1} = 55.49$, $SD_{MM-1} = 31.38$; *Jackie* (non-diagnostic), $M_{MM-1} = 73.00$, $SD_{MM-1} =$
 80 14.72; *Charlie* (non-diagnostic), $M_{MM-1} = 72.94$, $SD_{MM-1} = 14.79$. For the diagnostic
 81 descriptions, we observed significant variation depending on the description, $F(3,608) =$
 82 3.01, $p = .032$, partial $\eta^2 = 0.001$. When correcting for multiple comparisons, pairwise
 83 comparisons did not reveal significant differences between descriptions. We note that
 84 without correction, *Francis* appeared to be rated as more moral than both *Robin* ($p =$
 85 .012), and *Sam* ($p = .009$). For the non-diagnostic descriptions there was no significant
 86 difference in ratings depending on description, $t(211) = -0.09$, $p = .929$, $d = 0.01$.

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

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

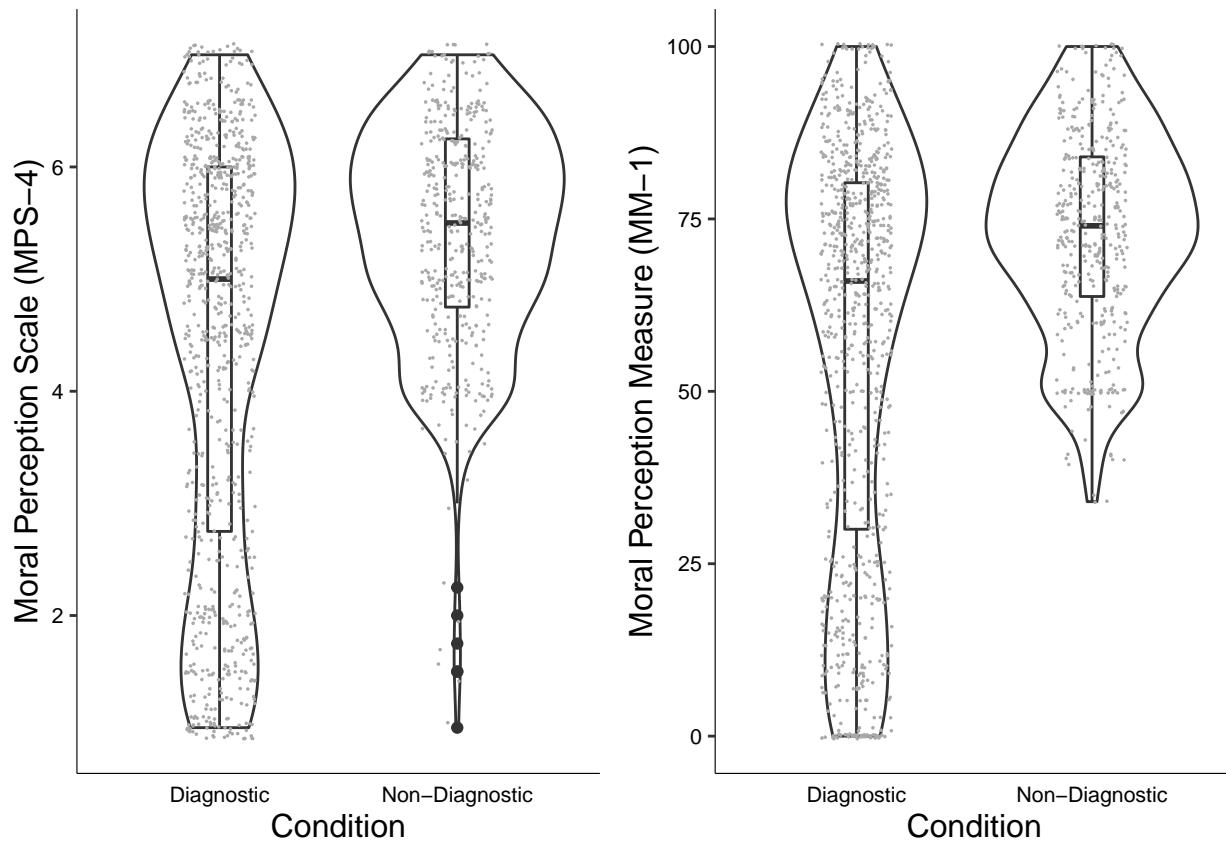


Figure 1. Pilot Study 1: Differences in moral perception depending on condition

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

Pilot Study 2

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

106 the diagnostic descriptions would be more extreme (more moral) than for the
107 non-diagnostic descriptions

108 **Pilot Study 2: Method**

109 **Pilot 2: Participants and design.** The pilot study was a within-subjects design.

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

114 A total sample of 245 (70 female, 175 male, 0 non-binary, 0 prefer not to say; $M_{age} =$
115 36.69, min = 18, max = 71, $SD = 9.57$) started the survey. Participants were recruited
116 from MTurk.

117 We removed participants who failed both manipulation checks ($n = 30$), leaving a
118 total sample of 215 participants (63 female, 152 male, 0 non-binary, 0 prefer not to say;
119 $M_{age} = 36.59$, min = 18, max = 71, $SD = 9.59$).

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

123 Moral character descriptions were developed by combining descriptions relating to
124 three different moral foundations, focusing on upholding the moral foundations (rather
125 than transgressions as in previous studies). We developed 4 descriptions of moral
126 characters that read as follows: (i) *Imagine a person named Sam. Throughout their life*
127 *they have been known to always help and care for others, treat everyone fairly and equally,*
128 *and show a strong sense of loyalty to others;* (ii) *Imagine a person named Robin.*
129 *Throughout their life they have been known to show compassion and empathy for others, act*
130 *with a sense of fairness and justice, and, never to break their word;* (iii) *Imagine a person*

131 named Francis. Throughout their life they have been known to uphold the standards of
 132 purity and decency, show respect for authority, and to always act honestly and fairly; (iv)
 133 Imagine a person named Alex. Throughout their life they have been known to protect and
 134 provide shelter to the weak and vulnerable, uphold the rights of others, and show respect for
 135 authority. We developed 2 descriptions of morally neutral characters that included
 136 information relating to physical appearance/attributes, hobbies/activities, and a color
 137 preference: (i) Imagine a person named Jackie. They have dark hair, go for a jog twice a
 138 week, and their favourite colour is blue; (ii) Imagine a person named Charlie. They have
 139 blue eyes, drink coffee in the morning, and their favourite colour is green.

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

142 Pilot 2: Results

143 The means and standard deviations for MPS-4 for each scenario are as follows: Sam
 144 (diagnostic), $M_{MPS-4} = 6.01$, $SD_{MPS-4} = 0.91$, Francis (diagnostic), $M_{MPS-4} = 5.89$,
 145 $SD_{MPS-4} = 0.95$, Alex (diagnostic), $M_{MPS-4} = 5.94$, $SD_{MPS-4} = 0.94$, Robin (diagnostic),
 146 $M_{MPS-4} = 5.93$, $SD_{MPS-4} = 0.92$, Jackie (non-diagnostic), $M_{MPS-4} = 5.60$, $SD_{MPS-4} = 0.99$,
 147 Charlie (non-diagnostic), $M_{MPS-4} = 5.53$, $SD_{MPS-4} = 1.08$. For the diagnostic descriptions,
 148 there was significant variation depending on the description, $F(3,613) = 2.91$, $p = .036$,
 149 partial $\eta^2 = 0.00$, Sam was viewed significantly more favorably than Francis ($p = .040$).
 150 For the non-diagnostic descriptions there was no significant difference in ratings depending
 151 on description, $t(214) = -1.79$, $p = .075$, $d = 0.12$.

152 The means and standard deviations for MM-1 for each scenario are as follows: Sam
 153 (diagnostic), $M_{MM-1} = 79.85$, $SD_{MM-1} = 15.44$; Francis (diagnostic), $M_{MM-1} = 78.30$,
 154 $SD_{MM-1} = 15.84$; Alex (diagnostic), $M_{MM-1} = 79.78$, $SD_{MM-1} = 15.71$; Robin (diagnostic),
 155 $M_{MM-1} = 79.46$, $SD_{MM-1} = 15.41$; Jackie (non-diagnostic), $M_{MM-1} = 73.44$, $SD_{MM-1} =$

¹⁵⁶ 15.83; *Charlie* (non-diagnostic), $M_{MM-1} = 73.07$, $SD_{MM-1} = 16.22$. For the diagnostic
¹⁵⁷ descriptions, we observed no significant variation depending on the description, $F(3,594) =$
¹⁵⁸ 1.45, $p = .231$, partial $\eta^2 = 0.002$. For the non-diagnostic descriptions there was no
¹⁵⁹ significant difference in ratings depending on description, $t(214) = -0.60$, $p = .552$, $d =$
¹⁶⁰ 0.04.

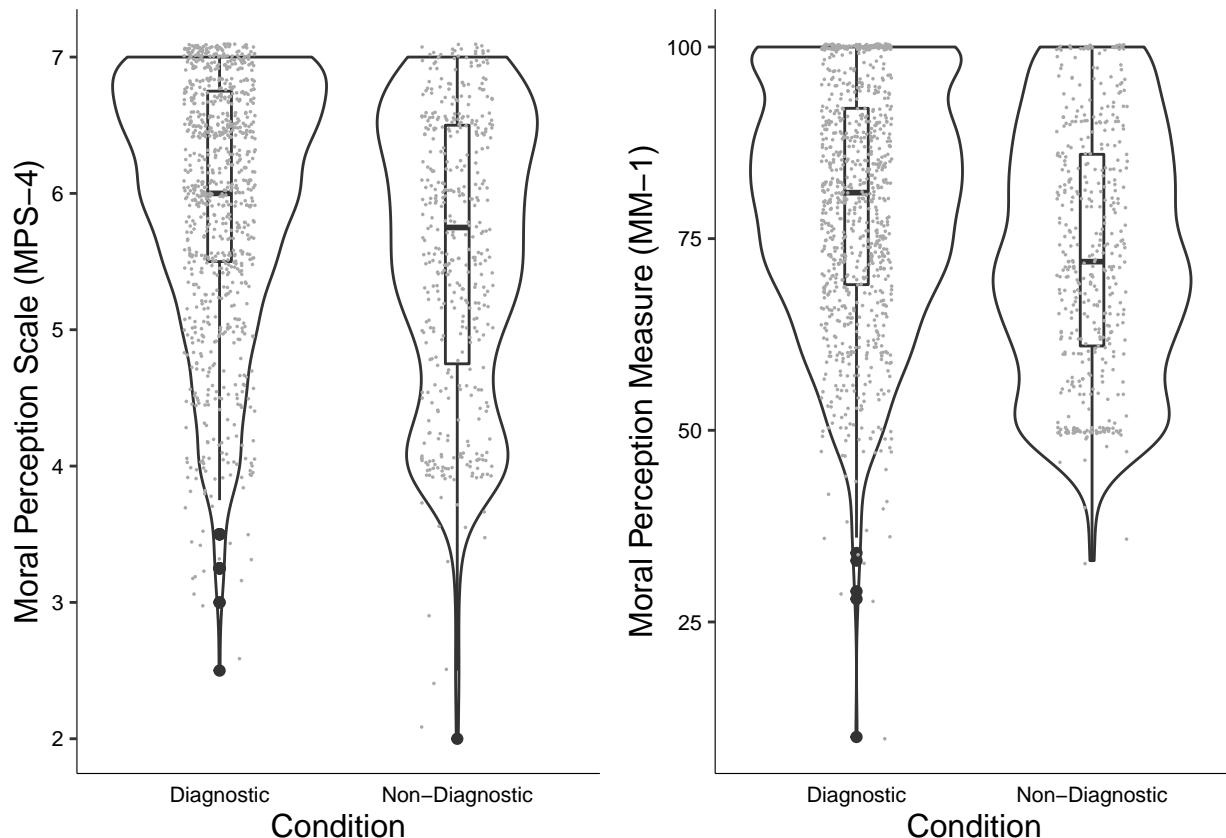


Figure 2. Pilot Study 2: Differences in moral perception depending on condition

¹⁶¹ We conducted a linear-mixed-effects model to test if condition influenced MPS-4
¹⁶² responses. Our outcome measure was MPS-4, our predictor variable was condition; we
¹⁶³ allowed intercepts and the effect of condition to vary across participants. Overall, the
¹⁶⁴ model significantly predicted participants responses, and provided a better fit for the data
¹⁶⁵ than the baseline model, $\chi^2(2) = 475.42$, $p < .001$. Condition was a significant predictor in
¹⁶⁶ the model $b = 0.19$, $t(214.35) = 6.53$, $p < .001$, with the diagnostic descriptions being

₁₆₇ rated as more moral than the non-diagnostic descriptions of immoral characters Figure 2.

₁₆₈ We conducted a linear-mixed-effects model to test if condition influenced MM-1
₁₆₉ responses. Our outcome measure was MM-1, our predictor variable was condition; we
₁₇₀ allowed intercepts and the effect of condition to vary across participants. Overall, the
₁₇₁ model significantly predicted participants responses, and provided a better fit for the data
₁₇₂ than the baseline model, $\chi^2(2) = 324.13, p < .001$. Condition was a significant predictor in
₁₇₃ the model $b = 3.04, t(214.90) = 6.02, p < .001$, with the diagnostic descriptions being
₁₇₄ rated as more moral than the non-diagnostic descriptions, see Figure 2.

175

Supplementary Materials

176 **Descriptions (Pilot Study 1 & Study 1)**

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

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

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

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

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

191 **Non-Diagnostic Descriptions.**

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

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

¹⁹⁶ **Descriptions (Pilot Study 2, Study 2 & Study 4)**

¹⁹⁷ **Diagnostic Descriptions.** Each moral description contains descriptive information
¹⁹⁸ relating to three different moral foundations as follows: *Sam*: care, fairness, loyalty; *Robin*:
¹⁹⁹ care, fairness, loyalty; *Francis*: purity, authority, fairness; *Alex*: care, fairness, authority.

²⁰⁰ **Sam.** Imagine a person named Sam. Throughout their life they have been known to
²⁰¹ always help and care for others, treat everyone fairly and equally, and show a strong sense
²⁰² of loyalty to others.

²⁰³ **Robin.** Imagine a person named Robin. Throughout their life they have been
²⁰⁴ known to show compassion and empathy for others, act with a sense of fairness and justice,
²⁰⁵ and, never to break their word.

²⁰⁶ **Francis.** Imagine a person named Francis. Throughout their life they have been
²⁰⁷ known to uphold the standards of purity and decency, show respect for authority, and to
²⁰⁸ always act honestly and fairly.

²⁰⁹ **Alex.** Imagine a person named Alex. Throughout their life they have been known
²¹⁰ to protect and provide shelter to the weak and vulnerable, uphold the rights of others, and
²¹¹ show respect for authority.

²¹² **Non-Diagnostic**

²¹³ **Jackie.** Imagine a person named Jackie. They have dark hair, go for a jog twice a
²¹⁴ week, and their favorite color is blue.

²¹⁵ **Charlie.** Imagine a person named Charlie. They have blue eyes, drink coffee in the
²¹⁶ morning, and their favorite color is green.

²¹⁷ **Descriptions (Study 3 & Study 5)**

²¹⁸ **Diagnostic Descriptions.**

²¹⁹ ***Sam (good).*** Imagine a person named Sam. Throughout their life they have been
²²⁰ known to always help and care for others, treat everyone fairly and equally, and show a
²²¹ strong sense of loyalty to others.

²²² ***Robin (good).*** Imagine a person named Robin. Throughout their life they have
²²³ been known to show compassion and empathy for others, act with a sense of fairness and
²²⁴ justice, and, never to break their word.

²²⁵ ***Alex (bad).*** Imagine a person named Alex. Throughout their life they have been
²²⁶ known to be cruel, act unfairly, and to betray their own group.

²²⁷ ***Francis (bad).*** Imagine a person named Francis. Throughout their life they have
²²⁸ been known to physically hurt others, treat some people differently to others, and show
²²⁹ lack of loyalty.

²³⁰ **Non Diagnostic Descriptions.** They have red hair, play tennis four times a
²³¹ month, and have one older sibling and one younger sibling.

²³² They are left-handed, drink tea in the morning, and have two older siblings and one
²³³ younger sibling.

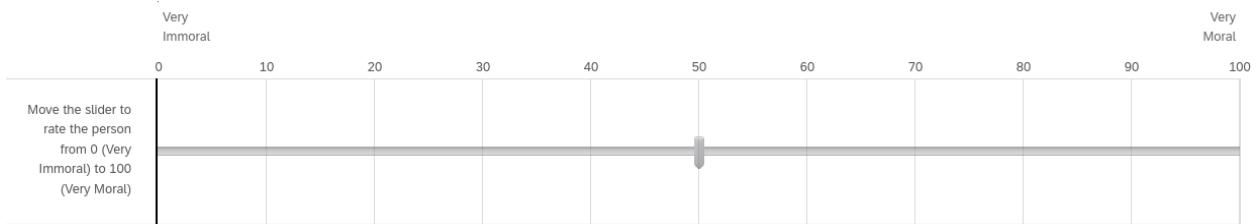
²³⁴ **Measures**

²³⁵ **Four-item Moral Perception Scale (MPS-4).** Please rate _____ along the
²³⁶ 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 3. Screenshot of the MPS-4 items as presented to participants

²³⁷ **Single-item Moral Perception Measure (MM-1).** Please rate _____
²³⁸ according to immoral or moral you view them:



²³⁹

240

Supplementary Analyses

241

Pilot Study 1

242 Pilot: 1: Differences Between Moral Descriptions

243 We developed a combined moral perception measure by calculating the mean of the

244 combined mean-centered scores for MPS-4 and MM-1, and mean-centering this result.

245 Below we report the analyses for this combined measure.

246 The standardized means and standard deviations for the combined measure for each

247 scenario are as follows: *Sam* (diagnostic), $M = -0.30$, $SD = 1.16$; *Francis* (diagnostic), M

248 = -0.22 , $SD = 1.06$; *Alex* (diagnostic), $M = -0.25$, $SD = 1.10$; *Robin* (diagnostic), $M =$

249 -0.31 , $SD = 1.19$; *Jackie* (non-diagnostic), $M = 0.36$, $SD = 0.55$; *Charlie* (non-diagnostic),

250 $M = 0.35$, $SD = 0.55$. For the moral descriptions, we observed significant variation

251 depending on the description, $F(3,602) = 2.67$, $p = .050$, partial $\eta^2 = 0.001$. When

252 correcting for multiple comparisons, pairwise comparisons did not reveal significant

253 differences between descriptions. We note that without correction, *Francis* appeared to be

254 rated as more moral than both *Robin* ($p = .022$), and *Sam* ($p = .021$). For the neutral

255 descriptions there was no significant difference in ratings depending on description, $t(211)$

256 = -0.46 , $p = .645$, $d = 0.03$.

257 Pilot 1: Testing Moral vs Neutral

258 We conducted a linear-mixed-effects model to test if condition influenced responses

259 on this combined measure. Overall, the model significantly predicted participants

260 responses, and provided a better fit for the data than the baseline model $\chi^2(2) = 1,035.36$,

261 $p < .001$, and condition was a significant predictor in the model $b = -0.31$, $t(210.99) =$

262 -8.74 , $p < .001$.



Figure 4. Pilot Study 1: Differences in combined measure depending on condition

263

Study 1

264

Again, we created a combined measure of moral perception from both DVs.

265

The means and standard deviations for the combined measure for each scenario are

266

as follows: *Sam*, $M = 0.02$, $SD = 0.89$, *Francis*, $M = 0.48$, $SD = 1.00$, *Alex*, $M = -0.21$, $SD = 0.92$, *Robin*, $M = -0.32$, $SD = 0.94$. There was significant variation depending on the description, $F(3,2255) = 269.01$, $p < .001$, partial $\eta^2 = 0.10$. *Francis* appeared to be rated as the most favorable, followed by *Sam*, then *Alex* and finally *Robin* as the least favorable (all $ps < .001$).

271

We conducted a linear-mixed-effects model to test if condition influenced moral perception. Our outcome measure was the combined moral perception measure, our

273 predictor variable was condition; we allowed intercepts and the effect of condition to vary
 274 across participants, and scenario was also included in the model. Overall, the model
 275 significantly predicted participants responses, and provided a better fit for the data than
 276 the baseline model, $\chi^2(8) = 762.31, p < .001$. Condition significantly influenced responses
 277 to the MPS-4, $F(1, 799.66) = 57.93, p < .001$; and was a significant predictor in the model
 278 when controlling for scenario, $b = -0.08, t(2,501.32) = -3.42, p < .001$, with the
 279 non-diagnostic descriptions being rated as more moral than the diagnostic (morally
 280 relevant) descriptions of immoral characters Figure 5.

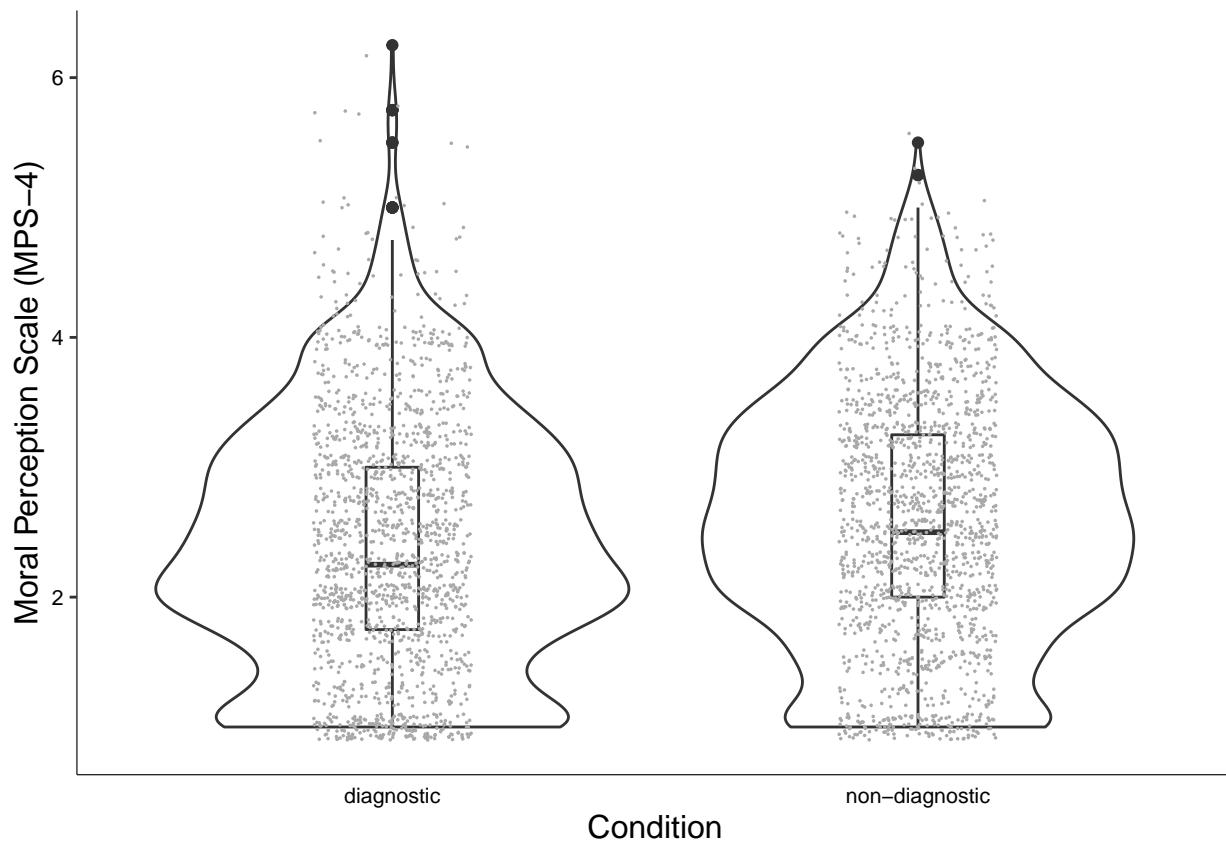


Figure 5. Study 1: Differences in combined measure depending on condition

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

282 We additionally conducted separate analyses for each scenario separately (for each
283 dependent measure MPS-4, MM-1 and the combined measure). The responses for each
284 scenario across each measure depending on condition are displayed in Figure 6.

285 For *Sam*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
286 = 2.70, SD = 0.82), than in the diagnostic condition (M = 2.42, SD = 0.87), $t(798.90)$ =
287 -4.66, p < .001, d = 0.33; MM-1 ratings were higher in the non-diagnostic condition (M =
288 26.55, SD = 16.41), than in the diagnostic condition (M = 21.50, SD = 15.59), $t(787.84)$ =
289 -4.45, p < .001, d = 0.32. For the combined measure ratings were also higher in the
290 non-diagnostic condition (M = 0.18, SD = 0.88), than in the diagnostic condition (M =
291 -0.13, SD = 0.88), $t(795.41)$ = -4.98, p < .001, d = 0.35.

292 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
293 condition (M = 2.16, SD = 0.90), than in the diagnostic condition (M = 2.09, SD = 0.92),
294 $t(793.94)$ = -1.09, p = .275, d = 0.08; MM-1 ratings were similar in the non-diagnostic
295 condition (M = 21.29, SD = 16.94), and in the diagnostic condition (M = 19.87, SD =
296 17.17), $t(794.97)$ = -1.18, p = .239, d = 0.08. For the combined measure ratings were also
297 similar in the non-diagnostic condition (M = -0.28, SD = 0.94), than in the diagnostic
298 condition (M = -0.36, SD = 0.94), $t(796.03)$ = -1.24, p = .217, d = 0.09.

299 For *Alex*, MPS-4 scores were significantly higher for the non-diagnostic condition (M
300 = 2.41, SD = 0.88), than in the diagnostic condition (M = 2.23, SD = 0.86), $t(796.97)$ =
301 -2.92, p = .004, d = 0.21; MM-1 ratings were higher in the non-diagnostic condition (M =
302 21.93, SD = 16.47), than in the diagnostic condition (M = 19.20, SD = 16.73), $t(798.89)$ =
303 -2.33, p = .020, d = 0.16. For the combined measure ratings were also higher in the
304 non-diagnostic condition (M = -0.12, SD = 0.92), than in the diagnostic condition (M =
305 -0.30, SD = 0.92), $t(798.40)$ = -2.82, p = .005, d = 0.20.

306 For *Francis*, MPS-4 scores were significantly higher for the non-diagnostic condition

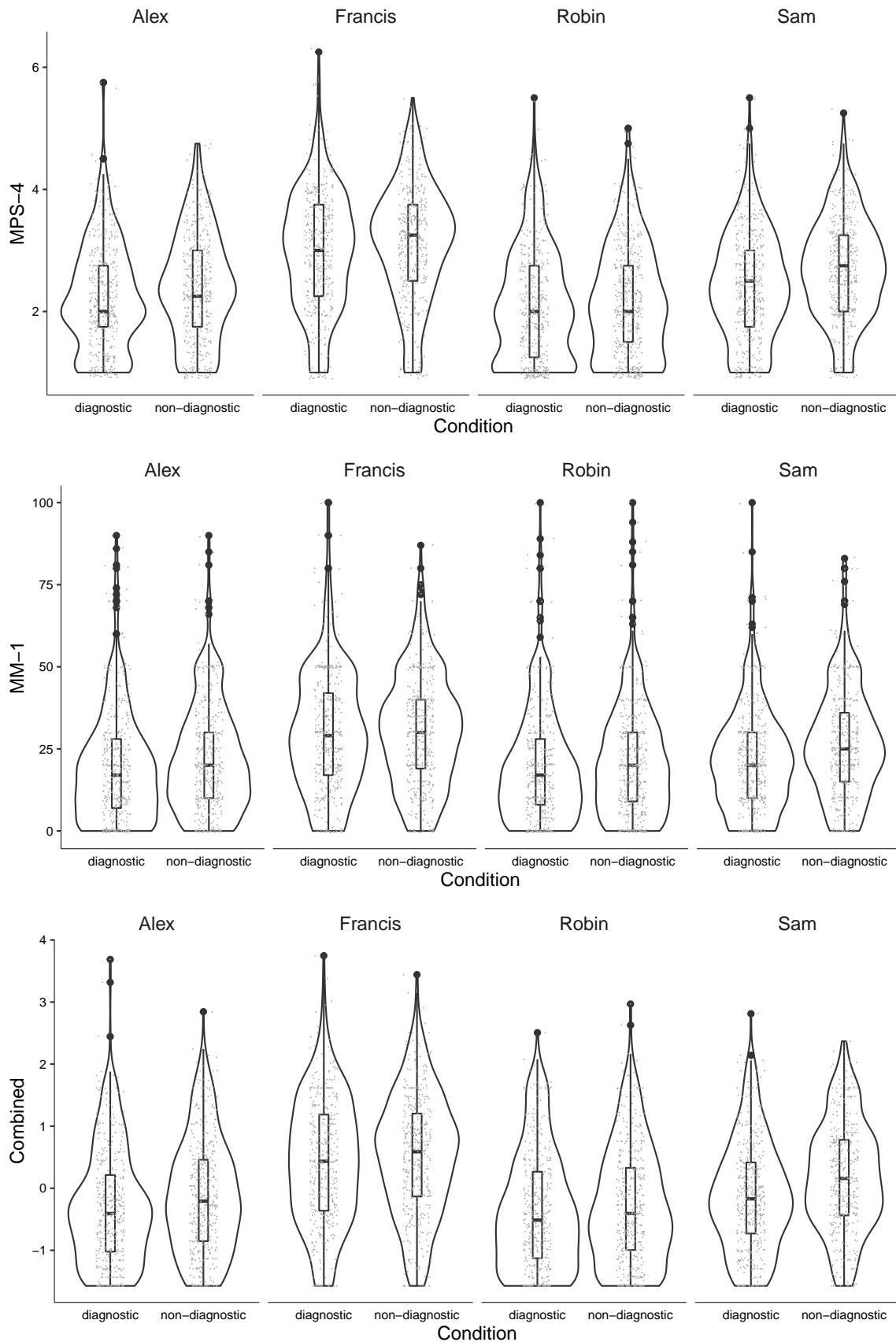


Figure 6. Study 1: Differences in moral perception for each description

307 ($M = 3.12$, $SD = 0.95$), than in the diagnostic condition ($M = 2.98$, $SD = 0.97$), $t(796.12)$
308 = -1.99, $p = .047$, $d = 0.14$; MM-1 ratings were not significantly different in the
309 non-diagnostic condition ($M = 30.38$, $SD = 17.17$), than in the diagnostic condition ($M =$
310 29.84, $SD = 18.56$), $t(788.61) = -0.43$, $p = .668$, $d = 0.03$. For the combined measure
311 ratings were also similar in the non-diagnostic condition ($M = 0.53$, $SD = 0.98$), and in the
312 diagnostic condition ($M = 0.44$, $SD = 1.02$), $t(794.36) = -1.29$, $p = .198$, $d = 0.09$.

313

Pilot Study 2

314 Pilot: 2: Differences Between Moral Descriptions

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

318 The standardized means and standard deviations for the combined measure for each
319 scenario are as follows: *Sam* (moral), $M = 0.21$, $SD = 0.91$; *Francis* (moral), $M = 0.10$, SD
320 = 0.96; *Alex* (moral), $M = 0.18$, $SD = 0.94$; *Robin* (moral), $M = 0.16$, $SD = 0.93$; *Jackie*
321 (neutral), $M = -0.24$, $SD = 1.01$; *Charlie* (neutral), $M = -0.30$, $SD = 1.07$. For the moral
322 descriptions, we observed significant variation depending on the description, $F(3,588) =$
323 2.90, $p = .039$, partial $\eta^2 = 0.002$. *Sam* was viewed significantly more favorably than *Francis*
324 ($p = .045$). For the neutral descriptions there was no significant difference in ratings
325 depending on description, $t(214) = -1.46$, $p = .147$, $d = 0.10$.

326 Pilot 2: Testing Moral vs Neutral

327 Overall, the model significantly predicted participants responses, and provided a
328 better fit for the data than the baseline model $\chi^2(2) = 564.98, p < .001$, and condition was
329 a significant predictor in the model $b = 0.22, t(214.32) = 6.60, p < .001$.

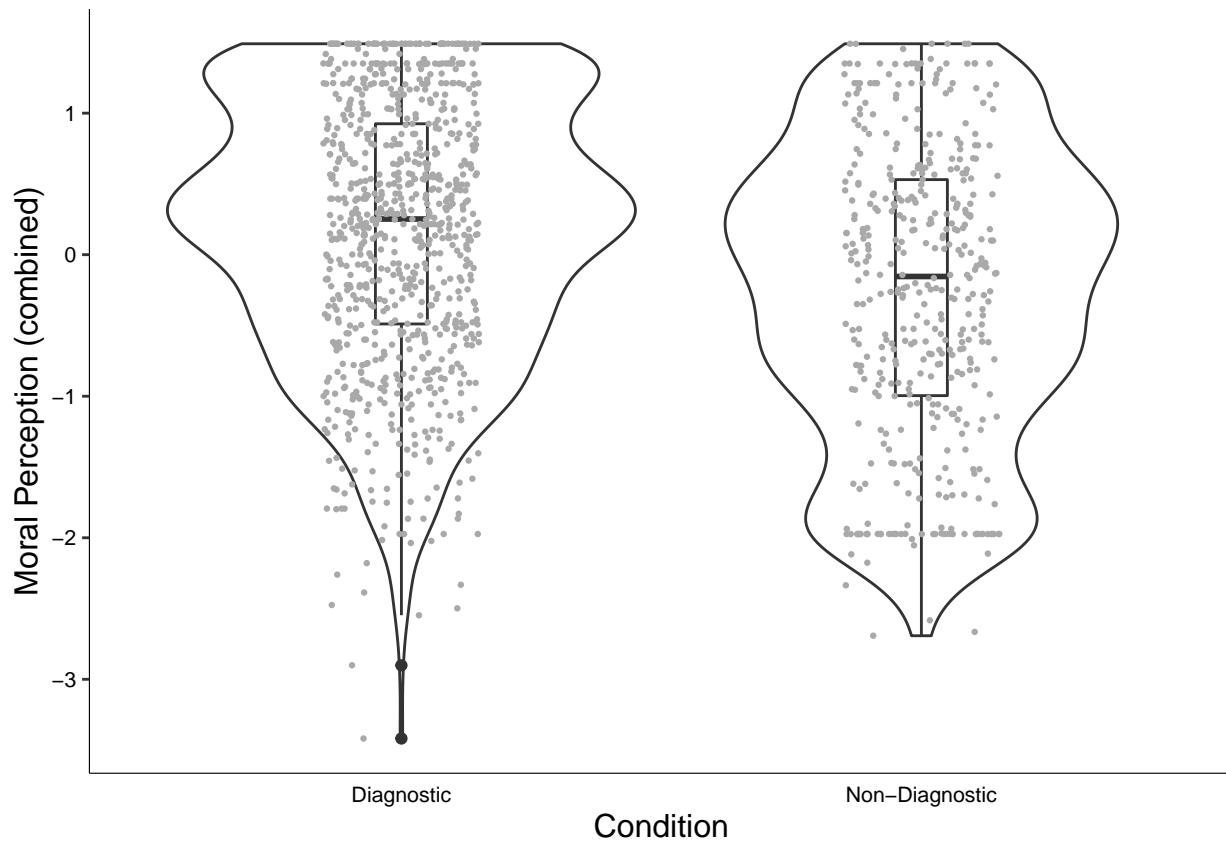


Figure 7. Pilot Study 1: Differences in combined measure depending on condition

330

Study 2

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

334 The means and standard deviations for the combined measure for each scenario are
335 as follows: *Sam*, $M = 0.07$, $SD = 0.97$, *Francis*, $M = -0.17$, $SD = 1.06$, *Alex*, $M = 0.09$, SD
336 = 1.02, *Robin*, $M = 0.07$, $SD = 0.96$. There was significant variation depending on the
337 description, $F(3,2335) = 48.01$, $p < .001$, partial $\eta^2 = 0.01$. *Francis* appeared to be rated
338 as the less favorable than all other characters (all $ps < .001$).

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

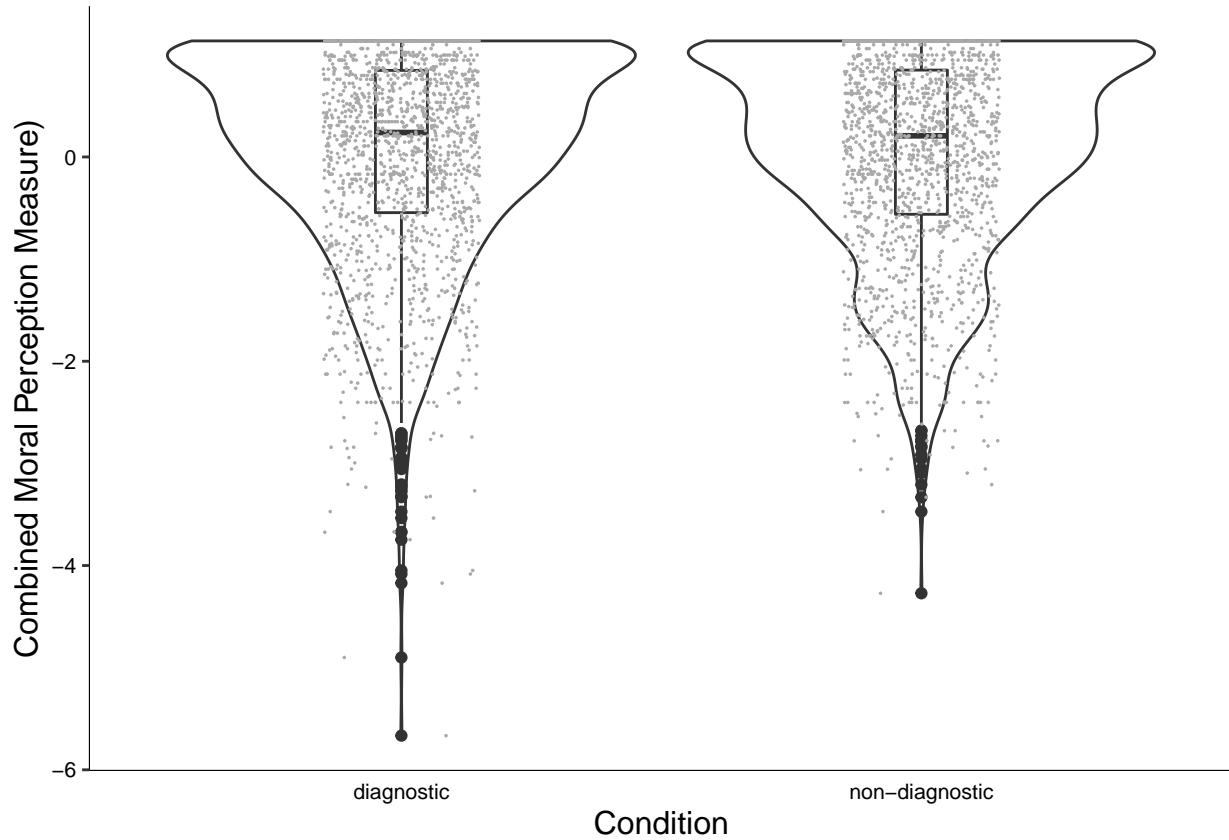


Figure 8. Study 2: Differences in combined measure depending on condition

³⁴⁷ Study 2: Differences between the Descriptions

³⁴⁸ Below we provide analyses of the effect of condition on responses to each scenario
³⁴⁹ individually. The responses for each scenario across each measure depending on condition
³⁵⁰ are displayed in Figure 9.

³⁵¹ For *Sam*, MPS-4 scores were not significantly different in the non-diagnostic condition
³⁵² ($M = 6.17$, $SD = 0.89$), than in the diagnostic condition ($M = 6.05$, $SD = 1.06$), $t(680.49)$
³⁵³ $= -1.71$, $p = .088$, $d = 0.12$; MM-1 ratings were similar in the non-diagnostic condition (M
³⁵⁴ $= 84.90$, $SD = 14.26$), than in the diagnostic condition ($M = 84.20$, $SD = 14.76$), $t(744.17)$
³⁵⁵ $= -0.69$, $p = .490$, $d = 0.05$. For the combined measure ratings were also similar in the
³⁵⁶ non-diagnostic condition ($M = 0.11$, $SD = 0.93$), than in the diagnostic condition ($M =$
³⁵⁷ 0.02 , $SD = 1.03$), $t(717.94) = -1.33$, $p = .183$, $d = 0.10$.

358 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
359 condition ($M = 6.08, SD = 1.00$), than in the diagnostic condition ($M = 6.13, SD = 0.98$),
360 $t(784.04) = 0.73, p = .463, d = 0.05$; MM-1 ratings were similar in the non-diagnostic
361 condition ($M = 84.12, SD = 14.37$), and in the diagnostic condition ($M = 85.98, SD =$
362 13.32), $t(800.09) = 1.92, p = .055, d = 0.13$. For the combined measure ratings were also
363 similar in the non-diagnostic condition ($M = 0.03, SD = 0.98$), than in the diagnostic
364 condition ($M = 0.13, SD = 0.95$), $t(788.76) = 1.46, p = .145, d = 0.10$.

365 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
366 condition ($M = 6.11, SD = 1.00$), than in the diagnostic condition ($M = 6.14, SD = 0.99$),
367 $t(737.60) = 0.32, p = .746, d = 0.02$; MM-1 ratings were similar in the non-diagnostic
368 condition ($M = 85.28, SD = 14.31$), than in the diagnostic condition ($M = 84.83, SD =$
369 15.51), $t(776.47) = -0.43, p = .668, d = 0.03$. For the combined measure ratings were also
370 similar in the non-diagnostic condition ($M = 0.09, SD = 0.98$), than in the diagnostic
371 condition ($M = 0.09, SD = 1.04$), $t(767.89) = -0.06, p = .952, d = 0.00$.

372 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
373 condition ($M = 5.82, SD = 1.05$), than in the diagnostic condition ($M = 5.90, SD = 1.08$),
374 $t(794.94) = 1.06, p = .290, d = 0.07$; MM-1 ratings were not significantly different in the
375 non-diagnostic condition ($M = 81.74, SD = 15.67$), than in the diagnostic condition ($M =$
376 $82.31, SD = 14.90$), $t(771.23) = 0.54, p = .591, d = 0.04$. For the combined measure
377 ratings were also similar in the non-diagnostic condition ($M = -0.20, SD = 1.08$), and in
378 the diagnostic condition ($M = -0.14, SD = 1.04$), $t(777.51) = 0.88, p = .379, d = 0.06$.

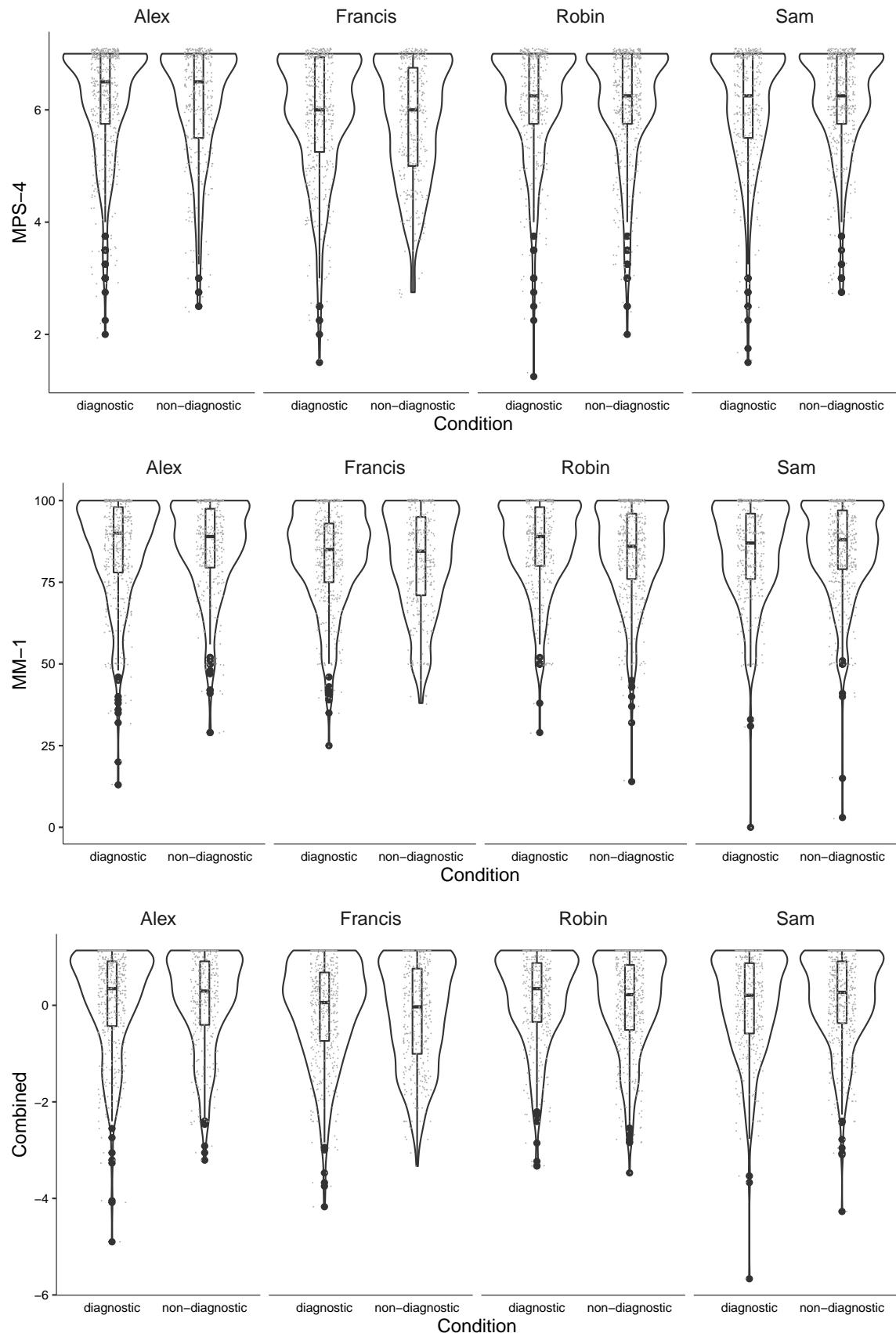


Figure 9. Study 2: Differences in moral perception for each description

379

Study 3

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

383 The means and standard deviations for the combined measure for each scenario are
384 as follows: *Sam*, $M = 0.45$, $SD = 0.52$, *Francis*, $M = -0.63$, $SD = 1.19$, *Alex*, $M = -0.66$,
385 $SD = 1.15$, *Robin*, $M = 0.43$, $SD = 0.52$. There was significant variation depending on the
386 description, $F(1,1027) = 473.77$, $p < .001$, partial $\eta^2 = 0.26$. Both the *good* characters
387 (*Robin* and *Sam*) were rated significantly more favorably than both the *bad* characters
388 (*Alex* and *Francis*; all $ps < .001$). There were no differences between *Robin* and *Sam* (*good*:
389 $p = .366$) or between *Alex* and *Francis* (*bad*; ($p = .648$)).

390 We conducted a linear-mixed-effects model to test if our predictors influenced
391 responses on the combined moral perception measure. Our outcome measure was the
392 combined moral perception measure, our predictor variables were condition and valence; we
393 allowed intercepts and the effects of condition and valence to vary across participants.
394 Overall, the model significantly predicted participants responses, and provided a better fit
395 for the data than the baseline model, $\chi^2(5) = 4,467.15$, $p < .001$. Condition significantly
396 influenced responses to the combined moral perception measure, $F(1, 873) = 16.65$, $p <$
397 $.001$ and was a significant predictor in the model when controlling for scenario, $b = -0.02$,
398 $t(873.00) = -4.08$, $p < .001$; valence significantly predicted responses, $F(1, 873) = 1,598.27$,
399 $p < .001$; and there was no significant condition \times valence interaction, $F(1, 873) = 0.03$, p
400 $= .867$, see Figure 10.

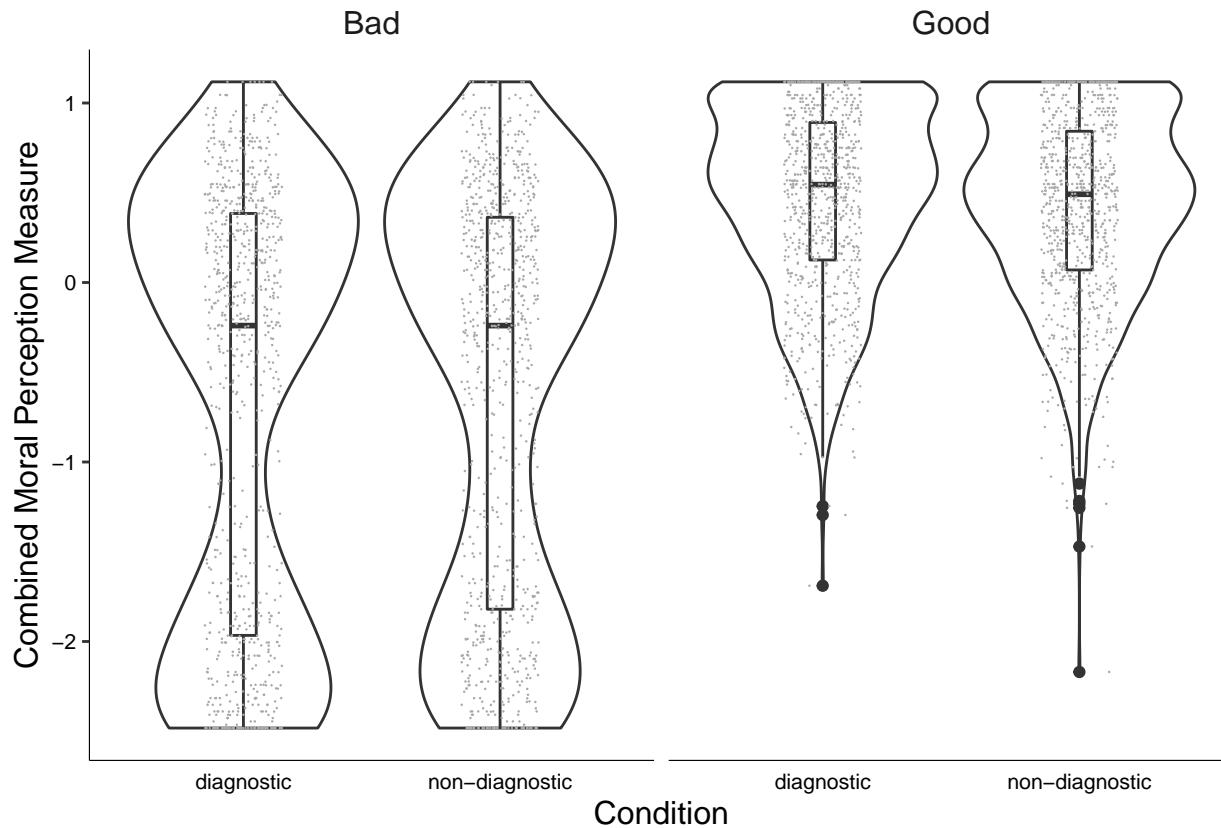


Figure 10. Study 3: Differences in the combined measure depending on condition

401 Study 3: Differences between the descriptions

402 Again, we conducted separate analyses to investigate of condition on responses to
 403 each scenario individually. The responses for each scenario across each measure depending
 404 on condition are displayed in Figure 11.

405 For *Sam*, MPS-4 scores were significantly lower in the non-diagnostic condition ($M =$
 406 5.81 , $SD = 1.09$), than in the diagnostic condition ($M = 5.98$, $SD = 0.97$), $t(859.15) =$
 407 2.46 , $p = .014$, $d = 0.17$; Similarly, MM-1 ratings were significantly lower in the
 408 non-diagnostic condition ($M = 79.64$, $SD = 15.68$), than in the diagnostic condition ($M =$
 409 82.37 , $SD = 14.67$), $t(867.08) = 2.66$, $p = .008$, $d = 0.18$. For the combined measure
 410 ratings were also lower in the non-diagnostic condition ($M = 0.39$, $SD = 0.54$), than in the
 411 diagnostic condition ($M = 0.50$, $SD = 0.50$), $t(863.14) = 2.85$, $p = .004$, $d = 0.19$.

412 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
413 condition ($M = 5.88, SD = 0.96$), than in the diagnostic condition ($M = 5.83, SD = 1.14$),
414 $t(844.53) = -0.77, p = .440, d = 0.05$; MM-1 ratings were similar in the non-diagnostic
415 condition ($M = 80.92, SD = 15.27$), and in the diagnostic condition ($M = 80.70, SD =$
416 15.07), $t(871.98) = -0.22, p = .828, d = 0.01$. For the combined measure ratings were also
417 similar in the non-diagnostic condition ($M = 0.44, SD = 0.51$), than in the diagnostic
418 condition ($M = 0.42, SD = 0.54$), $t(867.63) = -0.57, p = .569, d = 0.04$.

419 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
420 condition ($M = 4.08, SD = 1.96$), than in the diagnostic condition ($M = 3.97, SD = 2.11$),
421 $t(865.81) = -0.80, p = .421, d = 0.05$; MM-1 ratings were similar in the non-diagnostic
422 condition ($M = 52.19, SD = 31.29$), than in the diagnostic condition ($M = 49.58, SD =$
423 32.95), $t(868.76) = -1.20, p = .230, d = 0.08$. For the combined measure ratings were also
424 similar in the non-diagnostic condition ($M = -0.62, SD = 1.11$), than in the diagnostic
425 condition ($M = -0.70, SD = 1.19$), $t(867.67) = -1.04, p = .301, d = 0.07$.

426 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
427 condition ($M = 4.08, SD = 2.07$), than in the diagnostic condition ($M = 4.07, SD = 2.07$),
428 $t(871.94) = -0.09, p = .928, d = 0.01$; MM-1 ratings were not significantly different in the
429 non-diagnostic condition ($M = 51.56, SD = 32.68$), than in the diagnostic condition ($M =$
430 $51.42, SD = 33.70$), $t(871.59) = -0.06, p = .952, d = 0.00$. For the combined measure
431 ratings were also similar in the non-diagnostic condition ($M = -0.63, SD = 1.18$), and in
432 the diagnostic condition ($M = -0.64, SD = 1.20$), $t(871.88) = -0.08, p = .939, d = 0.01$.

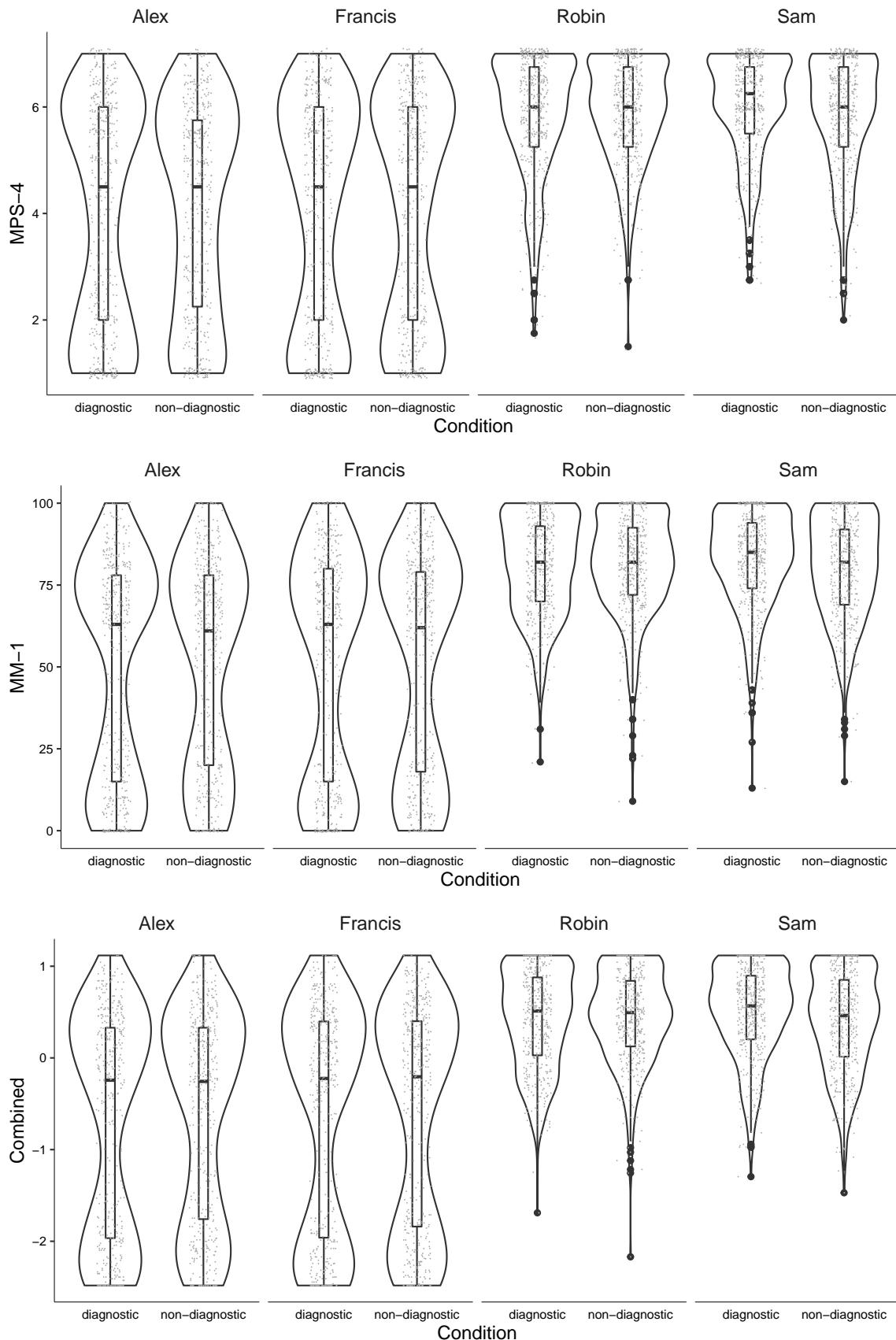


Figure 11. Study 3: Differences in moral perception for each description

433

Study 4

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

437 The means and standard deviations for the combined measure for each scenario are
438 as follows: *Sam*, $M = 0.03$, $SD = 1.02$, *Francis*, $M = -0.03$, $SD = 0.98$, *Alex*, $M = 0.02$, SD
439 = 1.04, *Robin*, $M = 0.04$, $SD = 1.01$. There was significant variation depending on the
440 description, $F(3,2493) = 4.32$, $p = .005$, partial $\eta^2 = 0.00$. *Francis* appeared to be rated as
441 the less favorable than all other characters (all $ps < .001$).

442 We conducted a linear-mixed-effects model to test if condition influenced moral
443 perception. Our outcome measure was the combined moral perception measure, our
444 predictor variable was condition; we allowed intercepts and the effect of condition to vary
445 across participants, and scenario was also included in the model. Overall, the model
446 significantly predicted participants responses, and provided a better fit for the data than
447 the baseline model, $\chi^2(8) = 42.42$, $p < .001$. Condition did not influence moral perception,
448 $F(1, 865.01) = 5.31$, $p = .021$; and was not a significant predictor in the model when
449 controlling for scenario, $b = -0.01$, $t(2,541.03) = -0.82$, $p = .410$, see Figure 8.

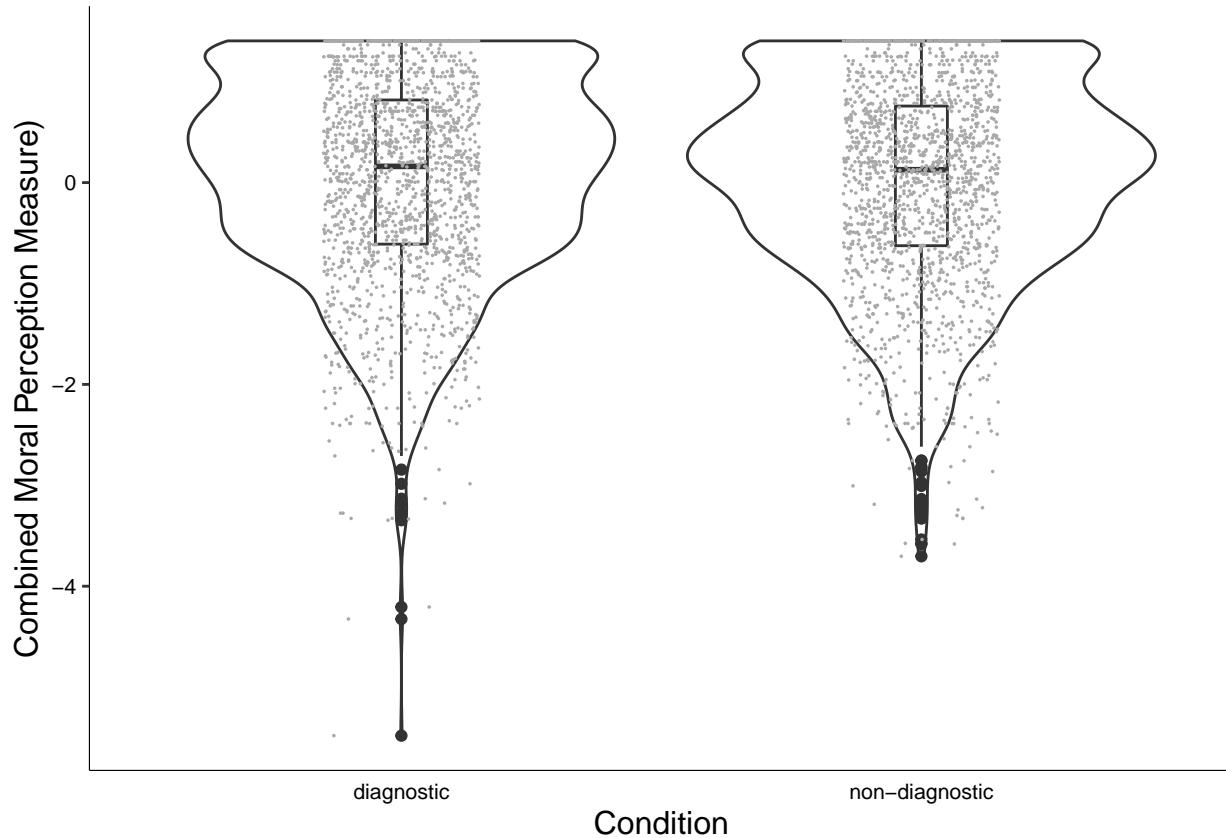


Figure 12. Study 4: Differences in combined measure depending on condition

450 Study 4: Differences between the Descriptions

451 As in previous studies, we additionally conducted separate analyses to investigate of

452 condition on responses to each scenario individually. The responses for each scenario across

453 each measure depending on condition are displayed in Figure 13.

454 For *Sam* (good), MPS-4 scores were significantly lower in the non-diagnostic

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

456 $t(810.53) = 1.97$, $p = .049$, $d = 0.14$; MM-1 ratings were similar in the non-diagnostic

457 condition ($M = 79.75$, $SD = 14.62$), than in the diagnostic condition ($M = 83.25$, $SD =$

458 13.30), $t(845.88) = 3.66$, $p < .001$, $d = 0.25$. For the combined measure ratings were also

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

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

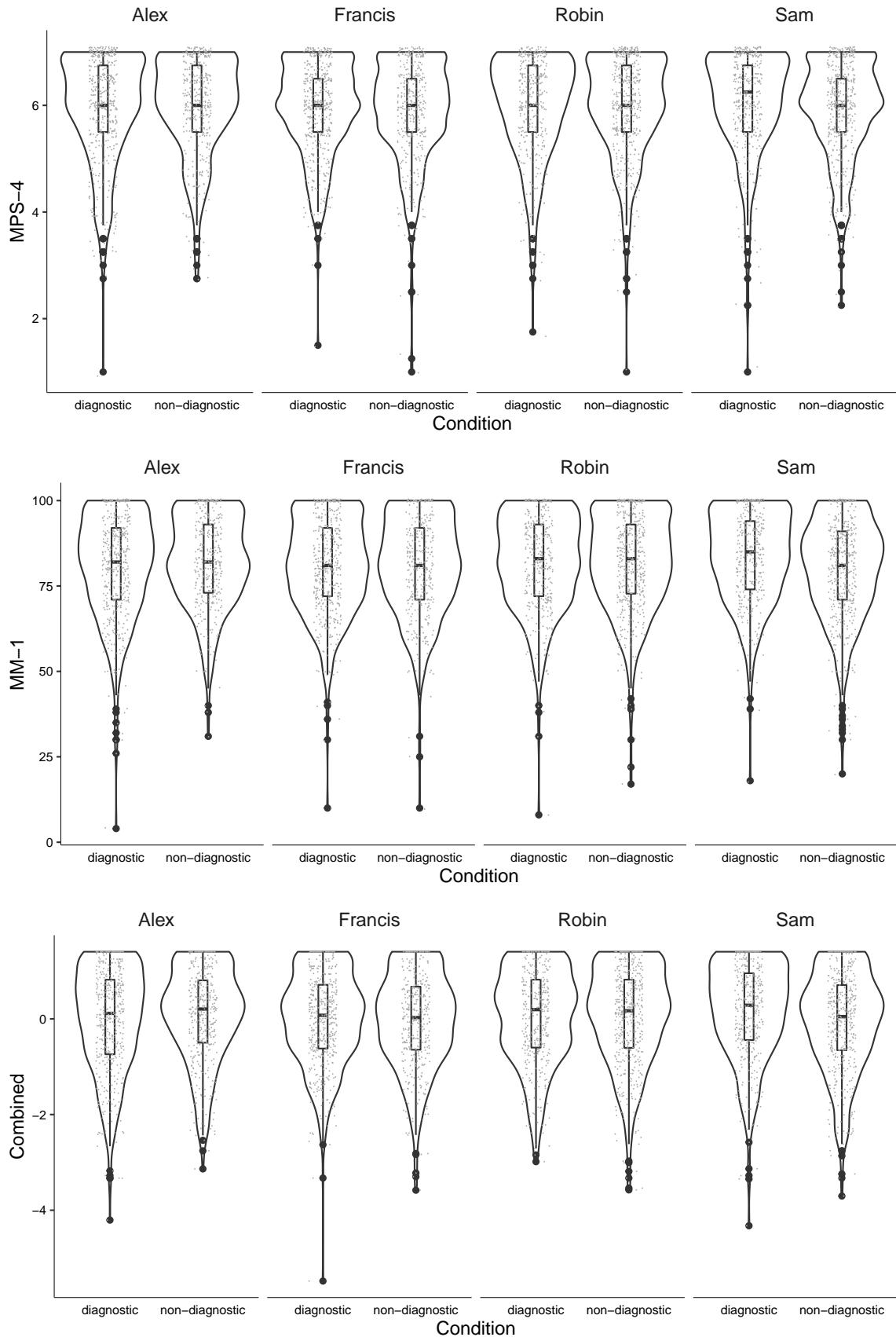


Figure 13. Study 4: Differences in moral perception for each description

461 For *Robin* (good), MPS-4 scores were significantly lower for the non-diagnostic
462 condition ($M = 5.95$, $SD = 0.93$), than in the diagnostic condition ($M = 5.94$, $SD = 0.95$),
463 $t(811.83) = -0.20$, $p = .841$, $d = 0.01$; MM-1 ratings were lower in the non-diagnostic
464 condition ($M = 81.62$, $SD = 14.28$), than in the diagnostic condition ($M = 81.64$, $SD =$
465 14.02), $t(824.54) = 0.02$, $p = .982$, $d = 0.00$. For the combined measure ratings were also
466 lower in the non-diagnostic condition ($M = 0.04$, $SD = 1.03$), than in the diagnostic
467 condition ($M = 0.04$, $SD = 0.99$), $t(828.47) = -0.10$, $p = .919$, $d = 0.01$.

468 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
469 condition ($M = 5.97$, $SD = 0.91$), than in the diagnostic condition ($M = 5.91$, $SD = 0.99$),
470 $t(845.29) = -0.91$, $p = .362$, $d = 0.06$; MM-1 ratings were similar in the non-diagnostic
471 condition ($M = 81.93$, $SD = 13.38$), than in the diagnostic condition ($M = 80.51$, $SD =$
472 15.21), $t(850.53) = -1.46$, $p = .145$, $d = 0.10$. For the combined measure ratings were also
473 similar in the non-diagnostic condition ($M = 0.07$, $SD = 0.98$), than in the diagnostic
474 condition ($M = -0.02$, $SD = 1.09$), $t(847.27) = -1.30$, $p = .192$, $d = 0.09$.

475 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
476 condition ($M = 5.87$, $SD = 0.95$), than in the diagnostic condition ($M = 5.91$, $SD = 0.87$),
477 $t(787.36) = 0.77$, $p = .443$, $d = 0.05$; MM-1 ratings were not significantly different in the
478 non-diagnostic condition ($M = 80.54$, $SD = 14.38$), than in the diagnostic condition ($M =$
479 80.75 , $SD = 13.99$), $t(809.63) = 0.21$, $p = .832$, $d = 0.01$. For the combined measure
480 ratings were also similar in the non-diagnostic condition ($M = -0.05$, $SD = 0.99$), and in
481 the diagnostic condition ($M = -0.01$, $SD = 0.98$), $t(814.30) = 0.55$, $p = .581$, $d = 0.04$.

482

Study 5

483 The means and standard deviations for the combined measure for each scenario are
484 as follows: *Sam*, $M = 84.52$, $SD = 15.49$; *Francis*, $M = 44.51$, $SD = 35.08$; *Alex*, $M =$
485 45.85 , $SD = 34.36$; *Robin*, $M = 85.15$, $SD = 14.61$. There was significant variation
486 depending on the description, $F(3,1746) = 351.55$, $p < .001$, partial $\eta^2 = 0.38$. Both the
487 *good* characters (*Robin* and *Sam*) were rated significantly more favorably than both the *bad*
488 characters (*Alex* and *Francis*; all $p < .001$). There were no differences between *Robin* and
489 *Sam* (*good*: $p = .963$) or between *Alex* and *Francis* (*bad*; ($p = .976$)).

490 We conducted a 2×2 between subjects ANOVA to test for an interaction between
491 valence and condition. Condition did not influence responses to the combined measure,
492 $F(1, 1746) = 8.42$, $p = .004$; valence significantly predicted responses, $F(1, 1746) = 964.98$,
493 $p < .001$; and there was no significant condition \times valence interaction, $F(1, 1746) = 0.04$, p
494 $= .841$.

495 For the *bad* characters, there was no significant difference in responses to the
496 combined measure between the diagnostic condition ($M = -0.81$, $SD = 1.15$) and the
497 non-diagnostic condition ($M = -0.70$, $SD = 1.07$) depending on condition, $t(834.36) =$
498 -1.23 , $p = .221$, $d = 0.09$.

499 For the *good* characters, there was a significant difference in responses to the
500 combined measure between the diagnostic condition ($M = 0.61$, $SD = 0.41$) and the
501 non-diagnostic condition ($M = 0.49$, $SD = 0.44$) depending on condition, $t(886.55) = 3.16$,
502 $p = .002$, $d = 0.28$.

503 **Study 5: Differences between the Descriptions**

504 Again, we conducted separate analyses to investigate of condition on responses to
505 each scenario individually. The responses for each scenario across each measure depending
506 on condition are displayed in Figure 14.

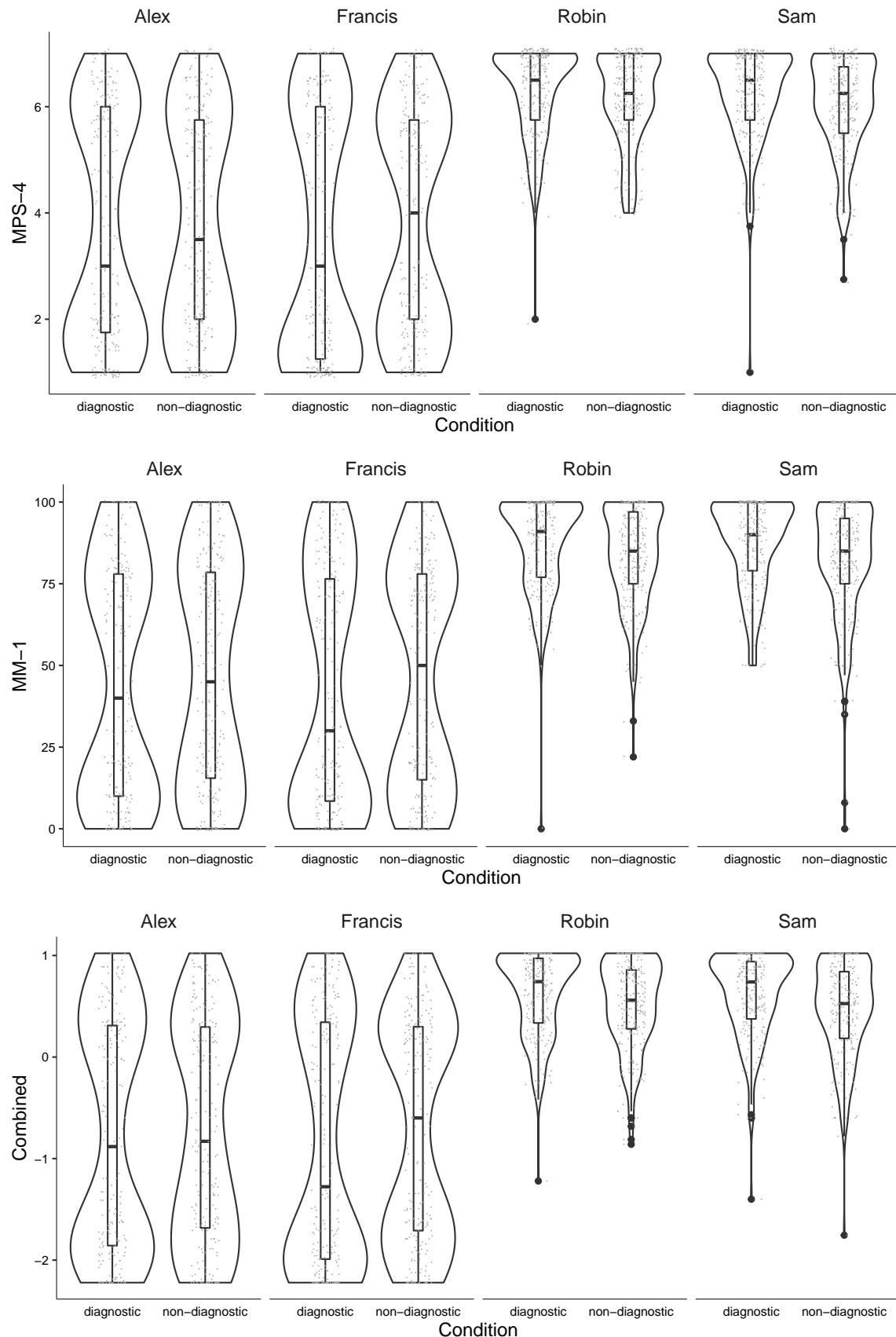


Figure 14. Study 5: Differences in moral perception for each description

507 For *Sam*, MPS-4 scores were not significantly different in the non-diagnostic condition
508 ($M = 5.87$, $SD = 0.95$), than in the diagnostic condition ($M = 5.91$, $SD = 0.87$), $t(886.55)$
509 = 3.16, $p = .002$, $d = 0.28$; MM-1 ratings were similar in the non-diagnostic condition (M
510 = 80.54, $SD = 14.38$), than in the diagnostic condition ($M = 80.75$, $SD = 13.99$), $t(809.63)$
511 = 0.21, $p = .832$, $d = 0.01$. For the combined measure ratings were also similar in the
512 non-diagnostic condition ($M = -0.05$, $SD = 0.99$), than in the diagnostic condition ($M =$
513 -0.01, $SD = 0.98$), $t(814.30) = 0.55$, $p = .581$, $d = 0.04$.

514 For *Robin*, MPS-4 scores were not significantly different for the non-diagnostic
515 condition ($M = 6.11$, $SD = 0.86$), than in the diagnostic condition ($M = 6.28$, $SD = 0.83$),
516 $t(448.03) = 2.09$, $p = .037$, $d = 0.20$; MM-1 ratings were similar in the non-diagnostic
517 condition ($M = 83.45$, $SD = 14.86$), and in the diagnostic condition ($M = 87.03$, $SD =$
518 14.14), $t(448.96) = 2.62$, $p = .009$, $d = 0.25$. For the combined measure ratings were also
519 similar in the non-diagnostic condition ($M = 0.51$, $SD = 0.43$), than in the diagnostic
520 condition ($M = 0.62$, $SD = 0.41$), $t(448.56) = 2.62$, $p = .009$, $d = 0.25$.

521 For *Alex*, MPS-4 scores were not significantly different for the non-diagnostic
522 condition ($M = 3.78$, $SD = 2.02$), than in the diagnostic condition ($M = 3.67$, $SD = 2.15$),
523 $t(406.27) = -0.55$, $p = .582$, $d = 0.05$; MM-1 ratings were similar in the non-diagnostic
524 condition ($M = 46.75$, $SD = 33.74$), than in the diagnostic condition ($M = 44.80$, $SD =$
525 35.13), $t(409.65) = -0.58$, $p = .560$, $d = 0.06$. For the combined measure ratings were also
526 similar in the non-diagnostic condition ($M = -0.71$, $SD = 1.06$), than in the diagnostic
527 condition ($M = -0.77$, $SD = 1.13$), $t(406.61) = -0.58$, $p = .562$, $d = 0.06$.

528 For *Francis*, MPS-4 scores were not significantly different for the non-diagnostic
529 condition ($M = 3.84$, $SD = 2.05$), than in the diagnostic condition ($M = 3.60$, $SD = 2.27$),
530 $t(424.52) = -1.17$, $p = .243$, $d = 0.11$; MM-1 ratings were not significantly different in the
531 non-diagnostic condition ($M = 46.97$, $SD = 34.05$), than in the diagnostic condition ($M =$
532 42.03, $SD = 35.99$), $t(428.22) = -1.47$, $p = .143$, $d = 0.14$. For the combined measure

⁵³³ ratings were also similar in the non-diagnostic condition ($M = -0.69$, $SD = 1.08$), and in
⁵³⁴ the diagnostic condition ($M = -0.84$, $SD = 1.18$), $t(425.90) = -1.35$, $p = .179$, $d = 0.13$.

535

References

- 536 Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., ... Eden,
537 A. (2020). Validating the extended character morality questionnaire. *Media Psychology*,
538 23(1), 107–130. <https://doi.org/10.1080/15213269.2019.1572523>
- 539 Walker, A. C., Turpin, M. H., Fugelsang, J. A., & Białek, M. (2021). Better the two devils
540 you know, than the one you don't: Predictability influences moral judgments of
541 immoral actors. *Journal of Experimental Social Psychology*, 97, 104220.
542 <https://doi.org/10.1016/j.jesp.2021.104220>