

The Fakeability of Personality Measurement with Graded
Paired Comparisons: *Supplement*

Item Analyses

Table 1

Item Statistics of Likert Items

Item	Desirability	Mean			SD		Skewness	
		Honest	Fake	Infl.	Honest	Fake	Honest	Fake
1	6.49	5.23	6.50	1.27	1.30	0.66	-0.67	-1.17
2	6.56	4.64	6.58	1.94	1.49	0.78	-0.43	-2.84
3	6.71	4.14	6.43	2.29	1.52	0.78	-0.17	-1.43
4	6.13	6.09	6.51	0.41	1.04	0.86	-1.53	-2.03
5	6.87	5.15	6.74	1.58	1.44	0.51	-0.72	-1.94
6	5.76	4.47	6.14	1.68	1.55	1.01	-0.18	-1.47
7	2.34	3.64	2.28	-1.36	1.55	1.89	0.33	1.57
8	1.66	3.82	1.74	-2.09	1.76	1.35	0.08	2.32
9	6.66	3.59	6.27	2.67	1.53	0.99	0.12	-1.63
10	2.43	4.08	2.31	-1.77	1.53	1.20	-0.01	0.98
11	5.73	3.89	5.93	2.04	1.56	1.10	0.20	-1.02
12	2.23	4.77	2.34	-2.42	1.35	1.25	-0.40	1.00
13	3.54	3.80	3.38	-0.42	1.43	1.40	0.19	-0.01
14	1.15	4.77	1.45	-3.32	1.55	1.00	-0.38	3.11
15	3.91	4.76	4.23	-0.53	1.59	1.74	-0.44	-0.09
16	5.17	3.63	5.42	1.79	1.55	1.49	0.17	-0.85
17	5.73	5.18	6.15	0.98	1.05	0.91	-0.59	-0.99
18	6.39	5.31	6.59	1.27	1.22	0.77	-0.88	-2.17
19	5.09	3.60	5.19	1.59	1.62	1.30	0.24	-0.38
20	3.86	4.10	3.87	-0.23	1.36	1.53	0.13	0.04
21	2.25	4.55	2.20	-2.36	1.47	1.22	-0.31	1.04
22	5.46	5.40	5.87	0.47	1.15	1.18	-0.89	-0.99
23	2.75	3.93	2.62	-1.32	1.49	1.40	0.12	1.04
24	1.19	3.74	1.30	-2.44	1.64	0.67	0.27	2.84
25	2.40	4.65	2.25	-2.40	1.70	1.44	-0.18	1.46
26	2.90	4.29	3.02	-1.27	1.17	1.49	-0.35	0.32
27	2.08	4.22	2.04	-2.19	1.58	1.06	-0.10	1.20
28	1.85	3.54	2.00	-1.54	1.60	1.22	0.21	1.23
29	6.48	4.80	6.55	1.75	1.51	0.66	-0.31	-1.25
30	5.40	5.05	5.49	0.44	1.20	1.20	-0.58	-0.65
31	6.30	3.40	5.97	2.56	1.48	1.32	0.31	-1.68
32	5.51	5.37	5.80	0.43	1.14	1.11	-0.64	-0.67
33	6.88	5.04	6.74	1.70	1.15	0.55	-0.49	-2.50
34	6.74	4.53	6.59	2.06	1.47	0.78	-0.44	-2.95
35	1.32	4.22	1.56	-2.66	1.54	1.19	-0.09	3.24
36	5.33	5.77	5.92	0.15	1.12	1.05	-1.03	-0.79
37	2.16	4.67	2.11	-2.56	1.38	1.15	-0.46	1.31

Table 1 continued

Item	Desirability	Mean			<i>SD</i>		Skewness	
		Honest	Fake	Infl.	Honest	Fake	Honest	Fake
38	6.57	4.49	6.58	2.09	1.45	0.72	-0.12	-2.13
39	6.90	5.57	6.86	1.29	1.00	0.44	-0.96	-4.19
40	2.99	5.22	3.00	-2.22	1.31	1.36	-0.72	0.36
41	5.61	3.95	5.97	2.02	1.61	1.29	0.04	-1.90
42	6.85	5.10	6.82	1.71	1.24	0.44	-0.45	-2.43
Mean								
all items	4.53	4.53	4.60	0.07	1.41	1.08	-0.29	-0.41

Note. Desirability ratings are based on responses from respondents in the Likert condition exclusively. The last row contains the column means of the table.

Table 2*Correlations of Likert Scale Item Statistics*

	1	2	3	4	5	6	7
1. Desirability	-						
2. Mean (honest)	.36*	-					
3. Mean (fake)	.99***	.40**	-				
4. Inflation	.95***	.07	.94***	-			
5. SD (honest)	-.43**	-.77***	-.46**	-.21	-		
6. SD (fake)	-.57***	-.46**	-.56***	-.43**	.35*	-	
7. Skew. (honest)	-.37*	-.97***	-.41**	-.09	.77***	.44**	-
8. Skew. (fake)	-.95***	-.37*	-.94***	-.88***	.48**	.54***	.41**

Note. Desirability = item desirability rated as rated by respondents of the Likert condition; Mean = item mean; SD = item standard deviation; honest = honest condition; faking = faking condition; Inflation = difference between honest and faking condition; SD = standard deviation; Skew. = item skewness. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3*Item Statistics of Graded Paired Comparisons*

GPC	Desirability			Mean			SD		Skewness	
	Item 1	Item 2	Diff.	Honest	Fake	Infl.	Honest	Fake	Honest	Fake
1	2.34	2.71	0.37	4.03	4.88	0.85	2.03	1.50	0.58	0.50
2	5.44	6.77	1.33	4.54	6.36	1.82	1.95	1.71	0.18	-0.18
3	6.42	2.33	-4.09	3.81	1.86	-1.94	1.84	1.56	0.52	2.24
4	5.98	5.60	-0.39	3.60	4.57	0.97	1.82	1.94	0.54	0.08
5	2.38	1.24	-1.13	5.31	4.05	-1.26	1.98	1.21	-0.21	-0.37
6	5.58	6.74	1.16	5.85	6.51	0.67	1.91	1.61	-0.23	-0.38
7	5.44	6.16	0.72	4.84	4.94	0.10	1.86	1.78	0.03	-0.11
8	2.34	4.97	2.63	3.94	6.81	2.87	1.98	1.71	0.47	-0.65
9	1.98	2.38	0.39	5.57	5.75	0.17	2.01	1.39	-0.19	-0.04
10	6.42	6.80	0.38	5.39	5.75	0.36	1.91	1.60	0.09	0.22
11	3.13	2.39	-0.74	4.02	4.11	0.10	1.64	1.23	0.32	-0.13
12	3.13	1.34	-1.80	4.00	3.45	-0.55	1.80	1.35	0.30	0.16
13	1.71	2.71	1.00	4.80	5.51	0.70	2.04	1.26	-0.18	0.61
14	1.71	6.80	5.10	5.86	8.39	2.53	1.99	1.19	-0.27	-2.26
15	1.98	3.81	1.83	6.15	6.66	0.52	2.07	1.72	-0.46	-0.28
16	4.97	6.77	1.80	6.16	7.34	1.18	1.72	1.54	-0.34	-0.78
17	6.45	1.34	-5.11	4.15	1.65	-2.50	2.00	1.23	0.32	2.63
18	3.77	2.23	-1.54	4.98	4.00	-0.98	1.97	1.55	0.12	0.31
19	5.50	1.24	-4.26	4.34	2.13	-2.21	1.92	1.29	0.19	1.28

Table 3 continued

GPC	Desirability			Mean			<i>SD</i>		Skewness	
	Item 1	Item 2	Diff.	Honest	Fake	Infl.	Honest	Fake	Honest	Fake
20	2.95	2.51	-0.44	5.33	4.56	-0.77	1.82	1.54	0.16	0.10
21	5.52	1.39	-4.14	4.09	2.08	-2.01	1.77	1.33	0.20	1.40
22	1.98	3.77	1.78	5.49	6.19	0.70	1.53	1.38	-0.35	0.27
23	6.16	6.30	0.14	4.84	5.16	0.32	2.04	1.64	0.22	0.05
24	5.02	6.42	1.41	5.95	7.08	1.13	1.59	1.54	-0.43	-0.44
25	3.46	1.71	-1.75	4.83	4.13	-0.71	1.63	1.22	0.19	0.04
26	6.66	6.45	-0.21	5.02	4.92	-0.09	1.80	1.79	0.07	0.16
27	2.39	6.33	3.94	5.64	8.12	2.47	2.04	1.25	-0.26	-1.78
28	5.98	6.30	0.32	3.96	4.80	0.85	1.85	1.61	0.37	0.07
29	3.13	2.39	-0.74	4.60	4.17	-0.43	1.50	1.23	0.01	-0.67
30	5.40	2.71	-2.69	3.29	2.63	-0.66	1.69	1.57	0.66	0.92
31	5.50	6.64	1.15	4.03	6.44	2.40	1.62	1.64	0.40	-0.25
32	5.50	3.81	-1.69	4.79	3.45	-1.34	1.83	1.73	0.12	0.78
33	5.52	4.97	-0.56	3.44	3.83	0.40	1.72	1.81	0.70	0.45
34	2.34	6.30	3.96	4.84	7.75	2.91	1.93	1.43	0.18	-1.53
35	3.46	6.66	3.20	5.33	7.78	2.45	1.90	1.40	-0.12	-1.49
36	5.50	2.38	-3.13	4.51	2.49	-2.02	1.83	1.40	0.22	1.12
37	6.30	6.74	0.44	5.24	6.17	0.93	1.73	1.48	0.02	0.12
38	5.50	2.51	-2.99	4.46	2.71	-1.75	1.87	1.53	0.31	1.07
39	5.42	1.71	-3.71	5.07	2.40	-2.68	2.18	1.61	-0.01	1.02
40	1.98	5.02	3.03	4.92	7.29	2.37	1.64	1.56	0.22	-0.70
41	6.16	4.97	-1.20	3.39	3.52	0.13	1.40	1.64	0.35	0.57
42	6.66	6.78	0.11	5.92	5.75	-0.17	1.89	1.46	-0.31	0.39
43	2.95	1.24	-1.71	5.72	4.01	-1.71	1.73	1.39	-0.24	-0.07
44	5.98	5.42	-0.56	3.30	3.48	0.19	1.47	1.60	0.21	0.37
45	6.64	2.51	-4.13	5.45	2.09	-3.36	1.99	1.42	-0.08	1.75
46	2.51	6.33	3.82	5.50	7.75	2.25	1.85	1.36	-0.17	-1.16
47	3.77	3.46	-0.31	5.10	4.47	-0.62	1.51	1.43	0.09	-0.09
48	5.02	1.39	-3.63	5.70	2.47	-3.24	1.72	1.65	-0.43	1.29
49	3.77	2.39	-1.37	5.09	3.87	-1.23	1.87	1.60	0.06	-0.08
50	3.77	5.02	1.25	4.64	6.27	1.63	1.59	2.00	0.26	-0.34
51	6.45	4.97	-1.48	4.03	3.05	-0.99	1.48	1.47	0.23	0.50
52	2.38	2.39	0.02	4.34	4.53	0.18	1.45	1.29	0.28	-0.10
53	5.40	2.34	-3.06	4.13	2.35	-1.78	1.79	1.45	0.28	1.23
54	2.95	6.29	3.34	4.65	7.43	2.78	1.54	1.71	0.27	-1.29
55	2.51	1.39	-1.12	5.12	4.40	-0.72	1.77	1.22	-0.28	-0.06
56	6.30	6.78	0.47	5.85	6.21	0.36	1.64	1.45	-0.04	0.24
57	5.40	6.78	1.38	4.52	5.80	1.29	1.77	1.54	0.38	-0.03
58	6.60	6.16	-0.43	6.01	4.09	-1.92	1.74	1.61	-0.32	0.22
59	2.39	4.97	2.57	4.71	6.80	2.09	1.92	1.57	0.01	-0.40
60	1.98	6.33	4.35	6.09	7.86	1.77	1.78	1.34	-0.22	-1.58

Table 3 continued

GPC	Desirability			Mean			<i>SD</i>		Skewness	
	Item 1	Item 2	Diff.	Honest	Fake	Infl.	Honest	Fake	Honest	Fake
61	5.52	2.39	-3.13	4.16	2.31	-1.85	1.75	1.42	0.48	1.40
62	3.13	5.60	2.46	4.80	7.65	2.85	2.08	1.50	0.05	-1.35
63	3.77	6.45	2.68	5.64	7.44	1.79	1.74	1.48	-0.01	-0.92
64	2.23	6.78	4.54	5.92	8.22	2.29	1.95	1.16	-0.36	-1.96
65	2.95	3.77	0.82	5.21	5.52	0.32	1.34	1.39	0.00	0.30
66	1.98	6.42	4.44	6.15	8.17	2.02	1.73	1.29	-0.41	-2.10
67	5.42	6.77	1.34	6.23	7.24	1.01	1.64	1.61	-0.41	-0.56
68	3.77	2.33	-1.44	4.85	4.06	-0.79	1.79	1.87	0.12	0.58
69	6.60	2.23	-4.37	5.31	1.83	-3.49	2.03	1.27	-0.01	2.18
70	2.95	6.30	3.35	6.03	7.67	1.64	1.70	1.51	-0.30	-1.33
71	3.46	3.13	-0.33	5.83	5.02	-0.80	1.64	1.42	-0.18	0.07
72	2.38	2.39	0.02	4.99	4.73	-0.26	1.50	1.05	0.19	-0.45
73	5.02	6.74	1.72	6.14	6.99	0.85	1.62	1.54	-0.34	-0.33
74	6.29	6.80	0.51	6.10	6.08	-0.02	1.55	1.62	-0.21	0.03
75	4.97	5.60	0.63	5.92	6.52	0.61	1.59	1.57	-0.24	-0.41
76	2.38	5.60	3.22	5.19	7.86	2.67	2.19	1.43	-0.13	-1.52
77	5.40	2.39	-3.01	3.40	2.13	-1.28	1.64	1.43	0.58	1.74
78	5.40	5.58	0.18	3.64	4.29	0.65	1.72	1.67	0.41	0.01
79	6.16	6.45	0.29	4.50	5.64	1.14	1.60	1.59	0.39	0.14
80	5.50	1.34	-4.16	4.20	2.07	-2.13	1.89	1.38	0.34	1.90
81	2.39	2.23	-0.16	5.42	5.14	-0.28	1.65	1.26	-0.12	0.25
82	1.98	2.71	0.72	5.39	5.13	-0.26	1.72	1.41	-0.17	0.38
83	3.81	1.39	-2.43	4.84	3.47	-1.37	1.91	1.42	-0.03	-0.10
84	2.95	1.98	-0.97	4.80	4.45	-0.36	1.48	1.20	0.26	0.15
85	2.39	2.39	0.00	4.43	4.74	0.31	1.63	1.04	0.02	-0.31
86	3.13	2.38	-0.76	4.56	4.63	0.07	1.44	1.14	-0.15	-0.34
87	2.39	3.81	1.42	5.30	6.14	0.84	1.86	1.55	-0.18	-0.07
88	2.39	2.51	0.12	4.97	4.96	-0.01	1.59	1.39	-0.07	0.24
89	2.95	4.97	2.02	4.85	6.29	1.44	1.53	1.62	0.07	-0.12
90	2.23	1.34	-0.90	4.70	4.21	-0.49	1.84	1.26	0.07	-0.01
91	6.29	5.58	-0.71	5.56	4.69	-0.87	1.59	1.86	-0.28	0.27
92	6.30	6.77	0.47	5.35	6.27	0.92	1.57	1.39	0.28	0.48
93	5.50	6.66	1.16	4.40	5.79	1.39	1.65	1.57	0.46	-0.01
94	5.50	6.74	1.24	5.06	6.21	1.15	1.80	1.58	0.19	-0.05
95	3.46	6.45	2.99	5.70	7.53	1.83	1.83	1.39	-0.28	-1.21
96	5.42	3.81	-1.61	5.77	4.56	-1.21	1.70	2.13	-0.18	0.05
97	2.95	3.46	0.51	5.12	4.96	-0.16	1.57	1.25	0.09	0.06
98	5.60	2.71	-2.89	4.60	2.58	-2.02	1.65	1.57	0.00	1.09
99	6.74	6.78	0.04	5.55	5.34	-0.21	1.57	1.25	-0.07	0.28
100	6.45	5.60	-0.85	4.89	4.40	-0.49	1.73	1.49	-0.18	0.05
101	2.95	3.13	0.18	6.03	5.03	-1.00	1.73	1.55	-0.39	-0.29

Table 3 continued

GPC	Desirability			Mean			<i>SD</i>		Skewness	
	Item 1	Item 2	Diff.	Honest	Fake	Infl.	Honest	Fake	Honest	Fake
102	5.98	2.23	-3.75	3.55	2.14	-1.41	1.67	1.43	0.53	1.73
103	6.16	5.60	-0.57	4.28	4.79	0.51	1.40	1.49	0.07	0.09
104	3.81	6.78	2.96	5.68	7.88	2.20	1.92	1.38	-0.12	-1.53
105	2.38	6.78	4.40	5.81	8.17	2.36	1.80	1.26	-0.23	-2.29
106	5.58	1.24	-4.34	5.15	2.10	-3.05	1.94	1.49	-0.05	1.91
107	6.60	2.71	-3.89	5.05	2.09	-2.95	1.75	1.47	-0.19	1.64
108	5.40	1.71	-3.69	3.57	1.88	-1.69	1.86	1.34	0.66	1.91
109	5.44	6.60	1.16	3.76	5.74	1.98	1.70	1.66	0.44	-0.04
110	6.64	1.24	-5.40	5.70	1.83	-3.86	1.88	1.45	-0.33	2.21
111	3.46	2.33	-1.13	4.98	4.51	-0.47	1.62	1.80	-0.03	0.77
112	3.77	6.80	3.04	5.79	7.87	2.08	1.73	1.33	-0.13	-1.72
113	3.46	6.77	3.30	5.81	8.01	2.20	1.74	1.29	-0.26	-1.99
114	6.74	6.77	0.03	5.04	5.09	0.05	1.38	1.37	-0.12	-0.04
115	6.29	2.33	-3.96	5.30	2.39	-2.91	1.84	1.86	-0.28	1.45
116	1.98	3.46	1.48	5.47	5.68	0.20	1.55	1.22	-0.02	0.04
117	2.39	6.30	3.91	5.41	8.10	2.68	2.07	1.32	-0.21	-2.30
118	6.78	6.77	-0.01	4.54	4.92	0.38	1.50	1.28	-0.25	-0.11
119	2.38	1.71	-0.67	4.56	4.38	-0.19	1.57	1.07	0.24	-0.46
Mean										
all GPCs	4.34	4.30	-0.04	4.97	5.05	0.08	1.75	1.47	0.03	0.04

Note. Diff. = difference between mean of item 1 and mean of item 2; Infl. = score inflation; Fake = faking condition. Desirability ratings are based on the single stimulus desirability ratings of respondents from the graded paired comparison condition exclusively. Note that items with negative factor loadings were not inverted for these calculations. The last row contains the column means of the table.

In order to provide information on our data on item level, we report central item statistics for Likert items in Table 1 and for GPCs in Table 3. Table 2 reports correlations between the variables of Table 1 (Likert items). Likewise, Table 4 reports correlations of the variables in Table 3 (GPC items). Apparently, the item means are very highly correlated with the average desirability rating of the items. We also analyzed the correlations between the responses in the Likert condition and the item desirability ratings of the same respondents. The mean correlations over the 42 Likert items was $r = 0.11$ in the honest and $r = 0.47$ in the faking condition. The mean standardized mean difference with honest *SD* as the denominator over all Likert items was $d = 1.14$.

In the Thurstonian IRT model with all 119 GPCs, the mean standardized threshold parameters were $M = -2.36$ ($SD = 0.37$) for threshold one, $M = -1.66$ ($SD = 0.34$) for threshold two, $M = -1.00$ ($SD = 0.38$) for threshold three, $M = -0.47$ ($SD = 0.42$)

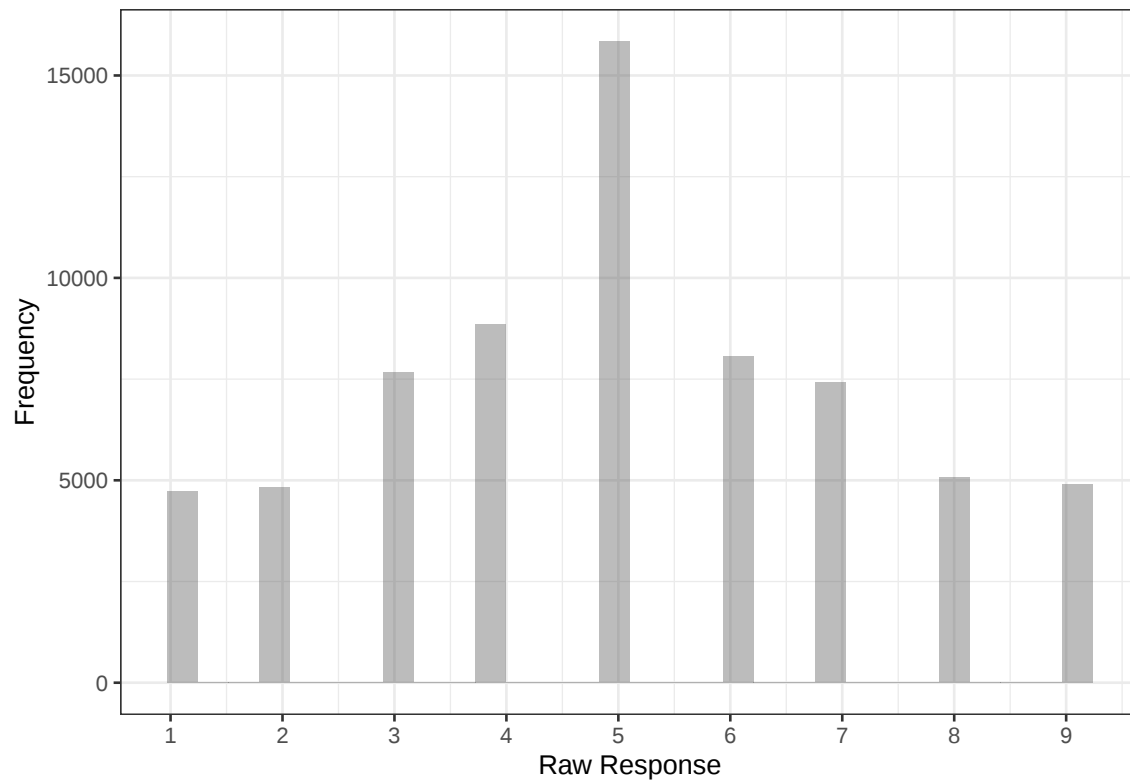


Figure 1

Histogram for Raw Responses on GPCs.

for threshold four, $M = 0.42$ ($SD = 0.40$) for threshold five, $M = 1.00$ ($SD = 0.40$) for threshold six, $M = 1.69$ ($SD = 0.34$) for threshold seven, and $M = 2.40$ ($SD = 0.36$) for threshold eight. For summary statistics of the standardized factor loadings, see the main article. We do not report summary statistics for the uniquenesses of the items because many of them were constrained to one for model identification and thus, summary statistics cannot be interpreted as usual.

Table 4*Correlations of Graded Paired Comparison Item Statistics*

	1	2	3	4	5	6	7	8	9
1. Des. item 1	-								
2. Des. item 2	.17	-							
3. Des. diff.	-.54***	.74***	-						
4. Mean (honest)	-.24**	.30**	.42***	-					
5. Mean (fake)	-.48***	.75***	.97***	.53***	-				
6. Inflation	-.45***	.74***	.94***	.15	.92***	-			
7. SD (honest)	-.10	-.07	.01	.12	.04	-.02	-		
8. SD (fake)	.39***	.24**	-.06	-.09	-.04	.00	.05	-	
9. Skew. (honest)	.23*	-.19*	-.32***	-.88***	-.40***	-.06	-.08	.14	-
10. Skew. (fake)	.58***	-.61***	-.92***	-.43***	-.90***	-.85***	-.02	.16	.38***

Note. Des. = Item desirability rated as rated by respondents of the GPC condition on a Likert scale; Des. diff. = difference of desirabilities (desirability of item 2 - desirability of item 1); Mean = item mean; SD = item standard deviation; honest = honest condition; faking = faking condition; Inflation = difference between honest and faking condition; SD = standard deviation; Skew. = item skewness. Note that many items have negative factor loadings. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Additional Information on the Analyses Reported in the Main Manuscript**Additional Descriptive Statistics**

Table 5 contains means and standard deviations of the main study variables for both the Likert and the GPC format. We see that honest scores have a mean of 0 and a SD of 1 due to the standardization procedure. The trait scores are inflated considerably from the honest to the faking condition in both formats (see means of faked trait scores). As was to be expected, the means match the estimates of the regression model for Hypothesis 1 in the main manuscript. Perceptions of trait desirabilities are very similar between the Likert and GPC condition. We present histograms for factor scores of Likert items in Figure 2 and of GPCs in Figure 3. In addition, Figure 4 shows the distribution of trait desirabilities. As intended by the instruction, all traits were perceived as highly desirable. For a discussion of the slightly lower desirability perceptions of agreeableness, see the main article.

Model Specification

In the following, we describe the regression models we used to test Hypotheses 1 to 4. To predict estimated trait scores, we used multivariate linear models that account for heterogeneous error variances. We denote the vector of estimated trait scores of person j as $\hat{\theta}_j = (\hat{\theta}_{j1}, \dots, \hat{\theta}_{jK})$ with elements $\hat{\theta}_{jk}$ denoting the estimated trait score of person

Table 5

Means and Standard Deviations of Trait Scores and Perceived Desirability

	Honest		Faking		Desirability	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Likert						
Emotional Stability	0.00	1.00	2.67	1.20	6.66	0.70
Extraversion	0.00	1.00	2.00	0.94	6.24	0.80
Agreeableness	0.00	1.00	0.96	1.23	5.63	1.20
Conscientiousness	0.00	1.00	3.10	1.33	6.84	0.53
GPC						
Emotional Stability	0.00	1.00	2.15	0.99	6.61	0.84
Extraversion	0.00	1.00	1.51	0.85	6.16	1.02
Agreeableness	0.00	1.00	1.15	1.27	5.85	1.12
Conscientiousness	0.00	1.00	2.04	1.03	6.82	0.62

Note. N = 290 in the Likert and N = 283 in the GPC format condition. Perceived trait desirability refers to the global trait rating on a scale from 1 to 7. Trait scores are standardized with the respective honest M and SD. Therefore, faked trait scores are equal to Cohen's d with the honest SD as denominator.

Table 6

Correlation Matrix for Honest (Lower Triangle) and Faked (Upper Triangle) Likert Scale Factor Scores with Perceived Trait Desirabilities

	1	2	3	4	5	6	7	8
1. Emotional Stability	-	.73***	.48***	.86***	.68***	.40***	.21***	.44***
2. Extraversion	.27***	-	.51***	.78***	.38***	.59***	.28***	.40***
3. Agreeableness	.10	.17**	-	.55***	.29***	.34***	.72***	.35***
4. Conscientiousness	.38***	.34***	.16**	-	.53***	.43***	.29***	.55***
5. Es Desirability	.14*	-.05	.10	.09	-	.37***	.16**	.43***
6. Ex Desirability	.01	.14*	.17**	.01	.37***	-	.29***	.37***
7. Ag Desirability	-.02	.12*	.14*	.10	.16**	.29***	-	.21***
8. Co Desirability	.07	.03	.14*	-.01	.43***	.37***	.21***	-

Note. Es = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; honest = honest condition; fake = faking condition. Upper triangle shows correlation with faked and lower triangle shows correlations with honest Likert scale factor scores and the perceived trait desirability. Perceived trait desirability is here calculated by the mean perceived item desirability for items measuring the respective trait. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 7

Correlation Matrix for Honest (Lower Triangle) and Faked (Upper Triangle) GPC Factor Scores with Perceived Trait Desirabilities

	1	2	3	4	5	6	7	8
1. Emotiona Stability	-	.82***	.64***	.82***	.52***	.36***	.32***	.37***
2. Extraversion	.59***	-	.59***	.74***	.41***	.56***	.38***	.34***
3. Agreeableness	.20***	.15*	-	.65***	.29***	.27***	.66***	.26***
4. Conscientiousness	.39***	.25***	.30***	-	.38***	.29***	.32***	.43***
5. Es Desirability	-.04	.06	.10	.03	-	.62***	.39***	.67***
6. Ex Desirability	-.10	.05	.10	.05	.62***	-	.48***	.59***
7. Ag Desirability	.00	.10	.24***	.11	.39***	.48***	-	.46***
8. Co Desirability	-.06	.08	.08	.01	.67***	.59***	.46***	-

Note. Es = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; honest = honest condition; fake = faking condition. Upper triangle shows correlation with faked and lower triangle shows correlations with honest GPC factor scores and the perceived trait desirability. Perceived trait desirability is here calculated by the mean perceived item desirability for items measuring the respective trait. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

j on trait k . The trait scores are assumed to be normally distributed with mean vector $\mu_j = (\mu_{j1}, \dots, \mu_{jK})$ and covariance matrix Σ_j

$$\hat{\theta}_j \sim \text{normal}(\mu_j, \Sigma_j). \quad (1)$$

The individual elements of the mean vector μ_{jk} are then predicted by a linear model which, in the most simple case, just contains an intercept b_{0k} and a slope b_{1k} for the dummy coded condition identifier (faking vs. honest)

$$\mu_{jk} = b_{0k} + b_{1k} \text{condition}_j. \quad (2)$$

The residual covariance matrix Σ_j is parameterized in terms of a vector of residual standard deviations $\sigma_j = (\sigma_{j1}, \dots, \sigma_{jK})$ and a residual correlation matrix Ω which is assumed be the same across all persons

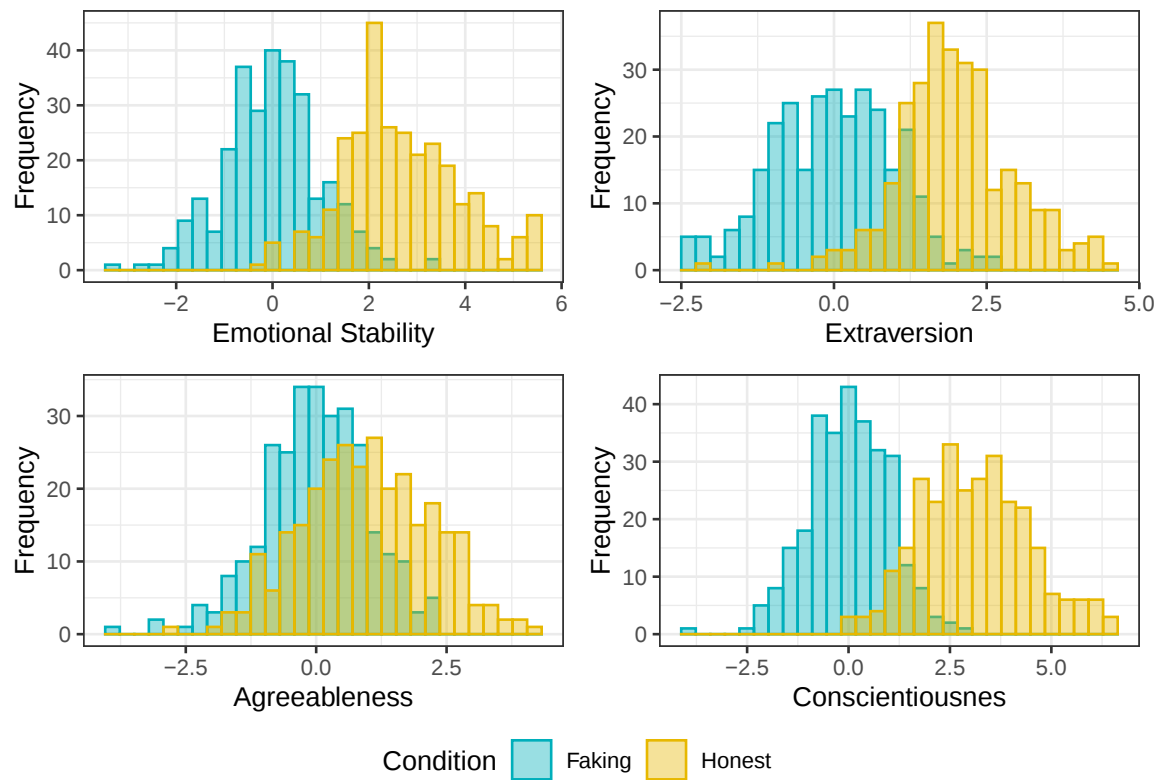
$$\Sigma_j = \text{diag}(\sigma_j) \Omega \text{diag}(\sigma_j). \quad (3)$$

To allow for modeling heterogeneous error variances not only across traits but also across conditions or other predictors, we predict σ_{jk} by a linear model as well. That is, in the most simple case, we specify

$$\sigma_{jk} = \exp(c_{0k} + c_{1k} \text{condition}_j) \quad (4)$$

with intercept c_{0k} and slope c_{1k} . Of course, it is possible to add more predictor terms both to the linear model of μ_{jk} and to the linear model of $\log(\sigma_{jk})$.

Control Variables

**Figure 2**

Histograms for Likert Scale Factor Scores.

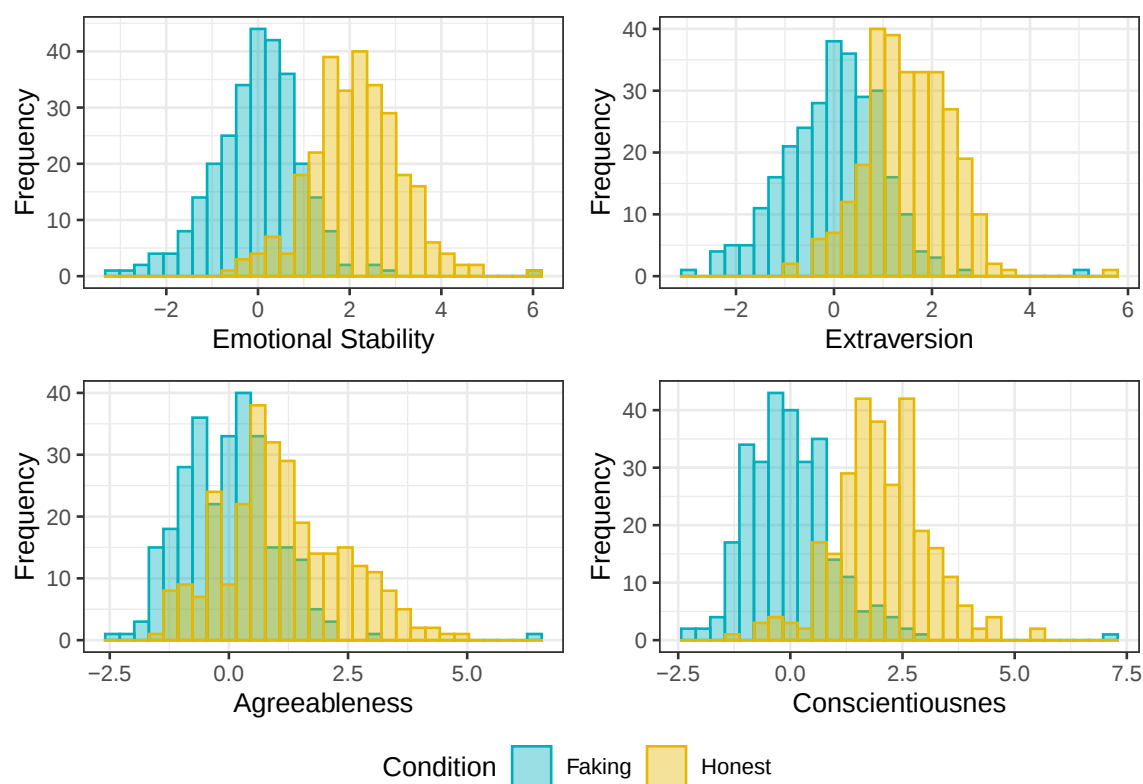


Figure 3
Histograms for Graded Paired Comparison Factor Scores.

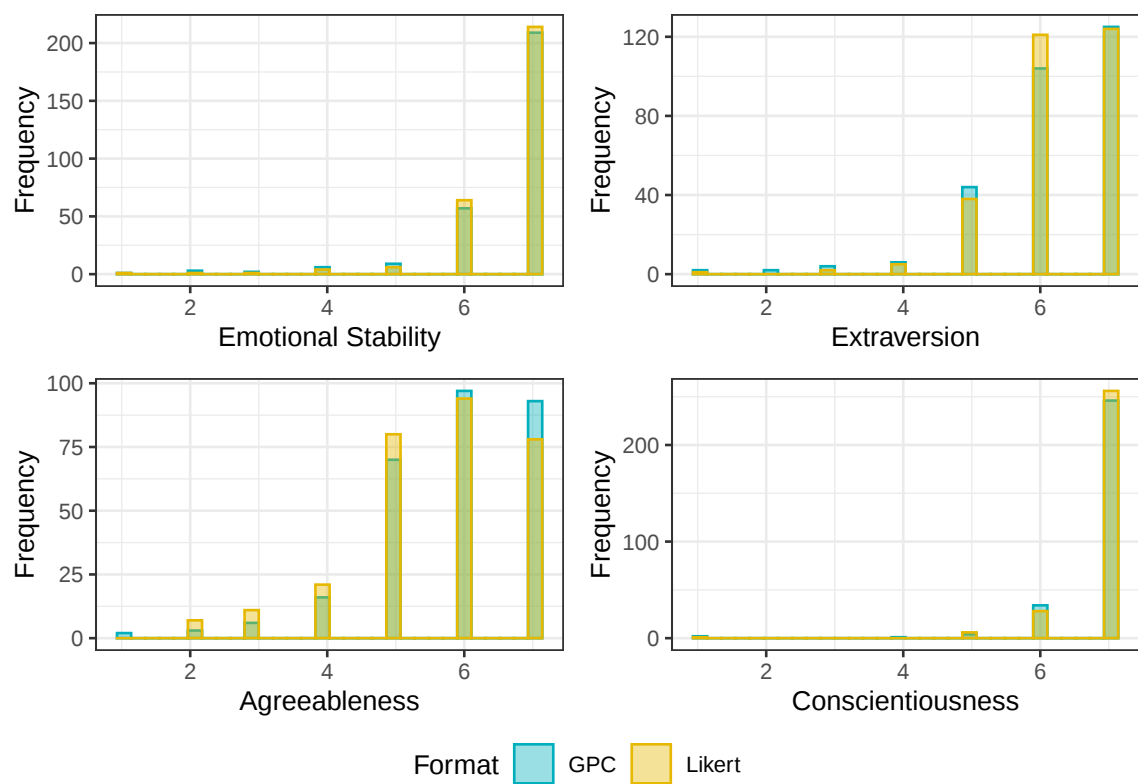


Figure 4
Histograms for Trait Desirabilities.

Table 8

Multivariate Regression Results With Honest Scores and Control Variables as a Predictors of Faked Scores

	Intercept	Honest	Gender	Age	Job	Duration	
Likert							
Emotional Stability	6.40 [3.53, 9.28]	0.19 [0.11, 0.26]	-0.15 [-0.49, 0.19]	-0.01 [-0.03, 0.01]	0.00 [-0.44, 0.45]	-0.47 [-0.89, -0.10]	
Extraversion	4.20 [2.07, 6.26]	0.17 [0.10, 0.23]	-0.12 [-0.37, 0.16]	-0.02 [-0.03, 0.00]	-0.02 [-0.35, 0.30]	-0.22 [-0.49, 0.07]	
Agreeableness	4.36 [1.55, 7.17]	0.26 [0.14, 0.39]	0.02 [-0.34, 0.38]	-0.03 [-0.05, -0.01]	-0.02 [-0.43, 0.40]	-0.34 [-0.77, 0.02]	
Conscientiousness	7.16 [4.05, 10.22]	0.17 [0.09, 0.24]	-0.30 [-0.66, 0.06]	-0.02 [-0.04, 0.00]	0.12 [-0.33, 0.59]	-0.46 [-0.89, -0.05]	
GPC							
Emotional Stability	0.30 [-1.83, 2.39]	0.14 [0.07, 0.21]	0.26 [-0.03, 0.54]	0.00 [-0.02, 0.02]	0.10 [-0.30, 0.50]	0.20 [-0.09, 0.47]	
Extraversion	0.64 [-1.51, 2.72]	0.04 [-0.03, 0.11]	0.07 [-0.20, 0.34]	0.00 [-0.02, 0.01]	0.28 [-0.07, 0.63]	0.09 [-0.19, 0.36]	
Agreeableness	-0.22 [-3.14, 2.73]	0.27 [0.14, 0.40]	-0.09 [-0.47, 0.28]	-0.03 [-0.06, -0.01]	-0.19 [-0.66, 0.30]	0.32 [-0.05, 0.68]	
Conscientiousness	0.54 [-1.74, 2.80]	0.03 [-0.05, 0.11]	-0.03 [-0.39, 0.31]	0.00 [-0.01, 0.02]	0.29 [-0.10, 0.67]	0.16 [-0.12, 0.45]	

Note. N = 290 in the Likert and N = 283 in the GPC format condition. Trait scores are standardized with the respective honest M and SD. For gender, male = 0, female = 1, diverse = NA; for Job, professional (i.e., unemployed, full-time, part-time) = 0, student = 1; Time = logarithmized response time for answering the entire questionnaire; GPC = graded paired comparisons. Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Results are based on separate models for each questionnaire format (Likert and GPC).

Full Regression Results for Hypothesis 4

Table 9 shows all regression coefficients for the model used to test Hypothesis 4 (the main manuscript reported only the three-way interaction terms). Although Hypothesis 4 was not supported, there is one result emerging from the joint regression model in Table 9 which highlights an advantage of the GPC format: The more individuals are motivated to fake, the higher their faked trait scores are on all traits over both questionnaire formats (see positive regression weights of faking intention). However the interaction of test format and faking intention is negative and significant for all traits but agreeableness. This means the effect becomes smaller when switching from the Likert to the GPC format. In other words, the individual faking intention is less associated with the faked trait scores when the GPC format is used, which argues for a somewhat higher faking resistance of this questionnaire format.

Table 9

Multivariate Multiple Regression With the Interaction of Honest Scores, Individual Faking Intention, and Questionnaire Format as Predictors for Faked Scores

	Intercept	Format	Honest	Intention	Hon*Form	Int*Form	Hon*Int	Hon*Int*Form
	Regression coefficients							
Es	2.65	-0.57	0.22	0.44	-0.11	-0.14	-0.12	-0.03
Ex	1.98	-0.52	0.20	0.36	-0.12	-0.13	-0.13	-0.03
Ag	0.97	0.14	0.27	0.14	0.01	-0.04	-0.09	-0.10
Co	3.07	-1.08	0.20	0.50	-0.19	-0.24	-0.14	-0.03
	95% CIs of regression coefficients							
Es	[2.52, 2.78]	[-0.73, -0.40]	[0.14, 0.29]	[0.30, 0.58]	[-0.21, -0.01]	[-0.32, 0.03]	[-0.19, -0.03]	[-0.13, 0.06]
Ex	[1.88, 2.09]	[-0.65, -0.38]	[0.13, 0.26]	[0.26, 0.46]	[-0.21, -0.03]	[-0.26, 0.01]	[-0.19, -0.07]	[-0.12, 0.06]
Ag	[0.83, 1.11]	[-0.05, 0.33]	[0.16, 0.39]	[0.01, 0.28]	[-0.16, 0.17]	[-0.22, 0.14]	[-0.20, 0.02]	[-0.25, 0.05]
Co	[2.92, 3.22]	[-1.27, -0.90]	[0.11, 0.28]	[0.34, 0.65]	[-0.30, -0.08]	[-0.43, -0.06]	[-0.22, -0.05]	[-0.14, 0.08]

Note. Format/Form = questionnaire format; Honest/Hon = honest score; Intention/Int = individual faking intention; ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Asterisk (*) = interaction terms of the variables involved; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Format is coded with Likert = 0, graded paired comparison = 1. Faking intention is z transformed.

Full Results for Quadratic Terms in the Analysis of Hypothesis 3

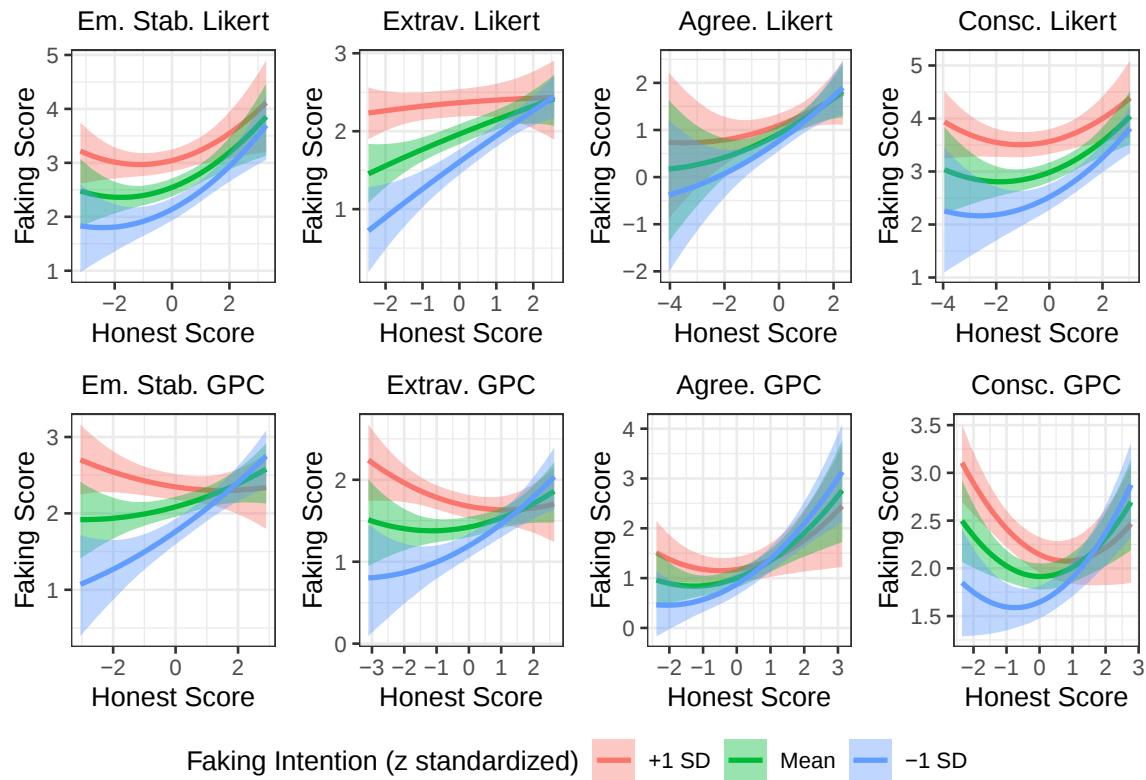


Figure 5

Interaction Plots for Hypothesis 3 Modeled with Linear and Quadratic Effects. Em. Stab. = Emotional stability; Extrav. = Extraversion; Agree. = Agreeableness; Consc. = Conscientiousness; GPC = Graded paired comparison condition. Shaded areas represent 95% credible intervals.

In the explorative results section, we report results for models that allowed for quadratic effects for both questionnaire formats. See Figure 5 for the interaction plots. For Likert scales (upper row) we see some small quadratic effects. Non-linear effects are more pronounced in GPCs but credibility intervals for all quadratic effects include zero.

Classical Test Theory Scoring for Likert scales

Because reliabilities and effect sizes are extraordinarily high for our IRT-based results in the main manuscript, we decided to check if these results hold in classical test theory analyses.

Table 10

Classical Test Theory Trait Scores and Effect Sizes for Likert Scales

	Honest		Faking		Cohen's d
	M	SD	M	SD	
Emotional Stability	3.58	1.08	5.89	0.78	2.13
Extraversion	4.15	1.16	5.88	0.68	1.48
Agreeableness	4.91	0.71	5.55	0.69	0.89
Conscientiousness	4.60	0.94	6.46	0.50	1.98

Note. $N = 290$ in the Likert condition. Cohen's d for difference between faked and honest scores with the honest SD as denominator.

Cronbach's α for Likert scales

In the manuscript, we report reliability estimates based on factor scores. Here, we report Cronbach's α based on the raw honest responses in the Likert questionnaire. Cronbach's α was 0.91 for emotional stability, 0.92 for extraversion, 0.78 for agreeableness, and 0.87 for conscientiousness.

Effect Sizes based on sum scores

We report standardized mean differences between honest and faking condition based on classical test theory sum scores in Table 10.

Alternative Hypothesis Test as Preregistered

t Tests for Hypothesis 1

The preregistered analysis method for Hypothesis 1 were t tests. As we argue in the manuscript, they are not the best way to analyze the data but we inform the reader about results in the following. We conducted t tests for dependent samples and heterogeneous variances to test the mean differences of factor scores between honest and faking condition. We report results separated for Likert scales and GPCs.

Likert

For emotional stability, we reach a mean difference of $M_d = 2.67$, 95% CI [2.51, 2.84], $t(289) = 31.69$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For extraversion, we reach a mean difference of $M_d = 2.00$, 95% CI [1.86, 2.14], $t(289) = 28.26$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For agreeableness, we reach a mean difference of $M_d = 0.96$, 95% CI [0.81, 1.12], $t(289) = 12.31$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For conscientiousness, we reach a mean difference of $M_d = 3.10$, 95% CI [2.93, 3.28], $t(289) = 35.17$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

The Bonferroni adjusted p values are $p < .001$ for all traits.

GPC

For emotional stability, we reach a mean difference of $M_d = 2.15$, 95% CI [1.99, 2.31], $t(282) = 26.26$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For extraversion, we reach a mean difference of $M_d = 1.51$, 95% CI [1.36, 1.66], $t(282) = 20.28$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For agreeableness, we reach a mean difference of $M_d = 1.15$, 95% CI [1.00, 1.29], $t(282) = 15.15$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

For conscientiousness, we reach a mean difference of $M_d = 2.04$, 95% CI [1.89, 2.19], $t(282) = 26.50$, $p < .001$, indicating that the mean honest factor score was significantly smaller in the honest than in the faking condition.

The Bonferroni adjusted p values are $p < .001$ for all traits.

Comparison of Correlations for Hypothesis 2

As an indicator of faking resistance, we investigated the association of honest and faking scores between the two formats for each trait (Hypothesis 2b). Within the manuscript, this is tested by the interaction term of Format*Honest Score. Here, we report an alternative approach which is based on the comparisons of honest-faking correlations of both questionnaire formats. To this end, we applied Fisher's-Z transformation to the original honest-faking correlations, calculated the Likert – GPC differences for each trait, and tested them against zero. The results are based on $n = 290$ in the Likert and $n = 283$ in the GPC condition.

For emotional stability, the correlation of honest and faked trait scores are $r = 0.16$ for the Likert and $r = 0.04$ for the GPC condition. The difference between both correlations is $\Delta r = 0.12$ (95% CI = $[-0.04, 0.28]$, $Fisher'sz = 1.44$, $p = 0.15$).

For extraversion, the correlation of honest and faked trait scores are $r = 0.23$ for the Likert and $r = 0.09$ for the GPC condition. The difference between both correlations is $\Delta r = 0.14$ (95% CI = $[-0.02, 0.30]$, $Fisher's z = 1.69$, $p = 0.09$).

For agreeableness, the correlation of honest and faked trait scores are $r = 0.30$ for the Likert and $r = 0.39$ for the GPC condition. The difference between both correlations is $\Delta r = -0.09$ (95% CI = $[-0.23, 0.05]$, $Fisher's z = -1.22$, $p = 0.22$).

For conscientiousness, the correlation of honest and faked trait scores are $r = 0.19$ for the Likert and $r = 0.18$ for the GPC condition. The difference between both correlations is $\Delta r = 0.01$ (95% CI = $[-0.15, 0.16]$, $Fisher's z = 0.07$, $p = 0.94$).

Exploratory Analysis with Only Equally-Keyed GPC Questionnaire

In the following, we will report results for Hypotheses 1 to 4 based on a GPC questionnaire with only equally keyed items. The data and results for the Likert questionnaire remain the same as in the main manuscript. The analysis procedure as well as the arrangement of tables and figures mirror those of the corresponding parts of the main manuscript. For explanations see the text of the main manuscript. One noteworthy difference is that the two analysis differ in terms of the standardization procedure. To enable comparisons between effect sizes of both GPC questionnaires (mixed and equally keyed) we used the mean and SD from the mixed-keyed questionnaire (honest condition) to standardize the factor scores of the equally keyed questionnaire in the faking condition. This was necessary to control for potential differences in variances between both questionnaires that would have differently affected the scaling procedure leading to effect sizes which cannot be compared directly. Honest scores were z -standardized for both GPC questionnaire versions.

Tables 11 shows the results for the regression of trait scores on the condition (honest vs. faking) for the Likert scales (upper rows) and the GPC questionnaire with equally keyed items (lower rows). The results for the Likert scales are equivalent to those from the main manuscript, since the data are the same (minimal differences are presumably due to the renewed Bayesian parameter estimation). Differences of effect sizes in the GPC condition (slopes) between equally keyed (Table 11) and mixed keyed (see corresponding table in the manuscript) are small and unsystematic. All slope CIs of both questionnaire versions overlap. Thus, the GPC questionnaire version based on only equally keyed pairs does not show less score inflation than the mixed-keyed version.

Results in Table 12 remain largely unchanged. For extraversion, the estimate of the factor *Format* still indicates lower faking effects for the GPC compared with the Likert questionnaire but the effect is no longer significant. Other changes with respect to the corresponding table for the mixed-keyed GPC questionnaire in the main manuscript are observed for agreeableness. For this trait, the faking effects are stronger in the GPC questionnaire with only equally keyed item pairs as compared with the Likert questionnaire. The estimate for the interaction term of honest score and questionnaire format for agreeableness is now on a comparable level as for the other traits.

Results in Table 13 show no relevant changes with respect to our hypotheses. The

Table 11

Equally-Keyed GPCs: Multivariate Regression Results With the Condition as a Predictor of Trait Scores

	Intercept (Honest)	Slope (Faking)
Likert		
Emotional Stability	0.00 [-0.14, 0.14]	2.67 [2.49, 2.86]
Extraversion	0.00 [-0.14, 0.13]	2.00 [1.84, 2.17]
Agreeableness	0.00 [-0.12, 0.12]	0.96 [0.79, 1.15]
Conscientiousness	0.00 [-0.14, 0.15]	3.10 [2.91, 3.30]
GPC		
Emotional Stability	-0.02 [-0.17, 0.12]	2.04 [1.86, 2.22]
Extraversion	-0.02 [-0.17, 0.12]	1.50 [1.34, 1.67]
Agreeableness	-0.02 [-0.17, 0.11]	1.21 [1.03, 1.40]
Conscientiousness	-0.02 [-0.17, 0.12]	1.92 [1.73, 2.11]

Note. Faking = factor score in faking condition; GPC = graded paired comparisons with only equally keyed item pairs. Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Results are based on separate models for each questionnaire format (Likert and GPC).

estimate for faking intention in the prediction of faked GPC scores for agreeableness significantly differs from zero now.

The pattern of results in Figure 6 mirror those from its mixed-keyed counterpart in the manuscript. One notable change is that for agreeableness measured with GPCs, honest and faking scores are now negatively correlated under high faking intention, as is the case with other traits measured with GPCs.

Table 12

Equally-Keyed GPCs: Multivariate Multiple Regression Results With the Interaction of Honest Scores and Format as Predictors for Faked Scores

	Intercept	Honest	Format	Honest*Format
Es	2.66 [2.49, 2.84]	0.21 [0.14, 0.28]	-0.65 [-0.85, -0.44]	-0.18 [-0.27, -0.08]
Ex	2.00 [1.87, 2.14]	0.18 [0.12, 0.24]	-0.52 [-0.68, -0.35]	-0.16 [-0.24, -0.08]
Ag	0.97 [0.80, 1.14]	0.24 [0.13, 0.35]	0.22 [0.02, 0.43]	-0.10 [-0.23, 0.04]
Co	3.10 [2.91, 3.30]	0.17 [0.09, 0.24]	-1.20 [-1.43, -0.98]	-0.16 [-0.25, -0.07]

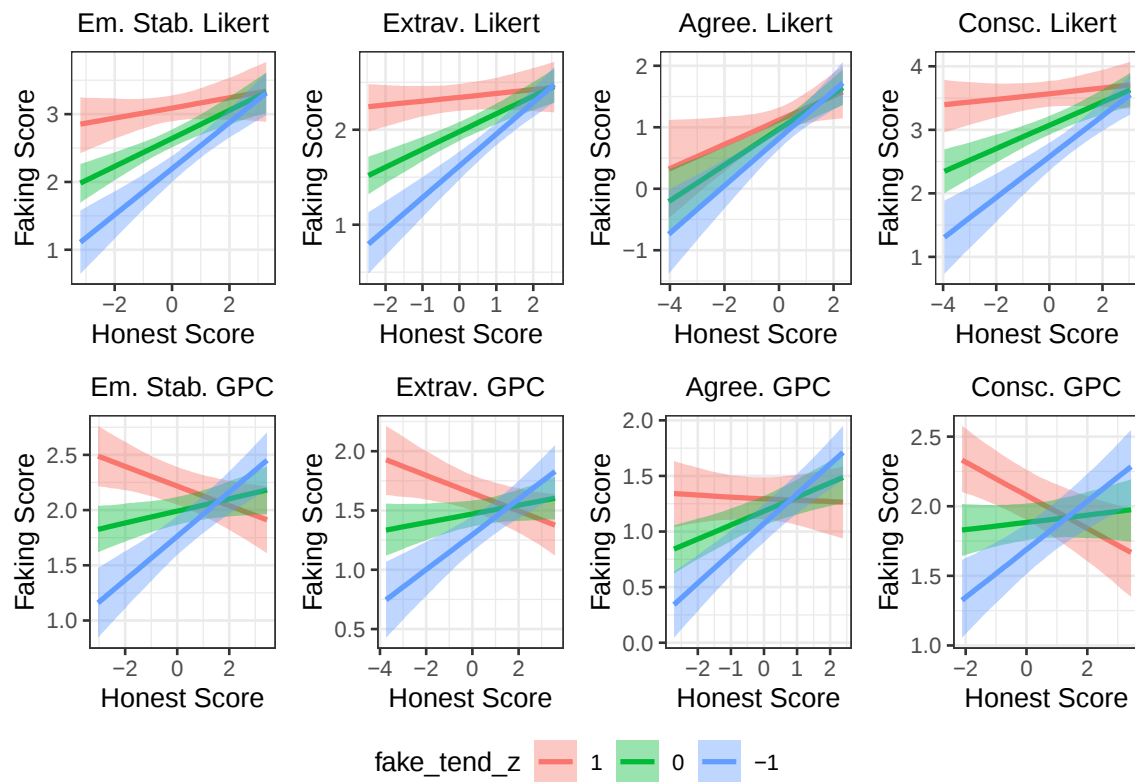
Note. ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Honest = honest scores; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Format is dummy coded with Likert = 0 and graded paired comparisons = 1. Honest*Format is the interaction effect of these two variables.

Table 13

Equally-Keyed GPCs: Multivariate Multiple Regression Results With the Interaction of Honest Scores and Individual Faking Intention as Predictors for Faked Scores

	Intercept	Honest	Faking Intention	Honest*Faking Intention
Likert				
Es	2.64 [2.51, 2.77]	0.21 [0.13, 0.29]	0.45 [0.30, 0.59]	-0.13 [-0.21, -0.05]
Ex	1.98 [1.88, 2.08]	0.19 [0.12, 0.25]	0.36 [0.26, 0.46]	-0.15 [-0.20, -0.09]
Ag	0.97 [0.83, 1.11]	0.29 [0.17, 0.41]	0.15 [0.02, 0.28]	-0.09 [-0.20, 0.02]
Co	3.07 [2.92, 3.22]	0.18 [0.11, 0.26]	0.50 [0.34, 0.66]	-0.14 [-0.21, -0.06]
GPC				
Es	1.99 [1.87, 2.12]	0.05 [0.00, 0.11]	0.23 [0.11, 0.35]	-0.14 [-0.19, -0.10]
Ex	1.47 [1.36, 1.58]	0.04 [-0.01, 0.08]	0.17 [0.07, 0.28]	-0.11 [-0.16, -0.07]
Ag	1.18 [1.05, 1.32]	0.13 [0.06, 0.19]	0.11 [-0.01, 0.23]	-0.14 [-0.20, -0.08]
Co	1.89 [1.76, 2.02]	0.03 [-0.03, 0.08]	0.19 [0.07, 0.32]	-0.15 [-0.20, -0.10]

Note. ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Honest = honest scores; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Faking intention is z transformed. Honest*Faking Intention is the interaction effect of these two variables.

**Figure 6**

Equally-Keyed GPCs: Interaction plots for Hypothesis 3. Em. Stab. = Emotional stability; Extrav. = Extraversion; Agree. = Agreeableness; Consc = Conscientiousness; GPC = Graded paired comparison condition.

Table 14

Equally-Keyed GPCs: Multivariate Multiple Regression With the Interaction of Honest Scores, Individual Faking Intention, and Questionnaire Format as Predictors for Faked Scores

	Intercept	Format	Hon Score	Intention	Hon*Form	Tend*Form	Hon*Tend	Hon*Tend*Form
	Regression coefficients							
Es	2.64	-0.65	0.21	0.44	-0.15	-0.22	-0.12	-0.02
Ex	1.98	-0.51	0.18	0.35	-0.14	-0.18	-0.14	0.03
Ag	0.97	0.22	0.24	0.14	-0.09	-0.03	-0.09	-0.06
Co	3.07	-1.18	0.18	0.49	-0.16	-0.30	-0.13	-0.01
	95% CIs of regression coefficients							
Es	[2.47, 2.81]	[-0.85, -0.45]	[0.13, 0.28]	[0.27, 0.61]	[-0.25, -0.06]	[-0.42, -0.02]	[-0.20, -0.05]	[-0.11, 0.07]
Ex	[1.85, 2.11]	[-0.67, -0.35]	[0.12, 0.24]	[0.23, 0.48]	[-0.22, -0.06]	[-0.33, -0.03]	[-0.19, -0.08]	[-0.05, 0.10]
Ag	[0.79, 1.13]	[0.01, 0.43]	[0.14, 0.35]	[-0.02, 0.30]	[-0.23, 0.04]	[-0.22, 0.16]	[-0.19, 0.02]	[-0.19, 0.06]
Co	[2.87, 3.26]	[-1.41, -0.97]	[0.10, 0.26]	[0.31, 0.68]	[-0.25, -0.07]	[-0.52, -0.09]	[-0.21, -0.06]	[-0.10, 0.08]

Note. Format = questionnaire format; Hon Score = honest score; Intention = individual faking intention; ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Asteriks (*) interaction terms of the variables involved; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Faking intention is z transformed.

In Table 14, we observe no relevant changes with consequences on our hypotheses.

Exploratory Analysis with GPC Questionnaire Matched for Social Desirability

In the following, we will report results for Hypotheses 1 to 4 based on a GPC questionnaire that is based on the 50 GPCs with the lowest difference in social desirability. The data and results for the Likert questionnaire remain the same as in the main manuscript. The analysis procedure as well as the arrangement of tables and figures mirror those of the corresponding parts of the main manuscript. For explanations see the text of the main manuscript. One noteworthy difference is that the two analysis differ in terms of the standardization procedure. To enable comparisons between effect sizes of both GPC questionnaires (full length and 50 pairs with lowest desirability difference) we used the mean and SD from the GPC questionnaire of full length (honest condition) to standardize the factor scores of the questionnaire matched for social desirability in the faking condition. This was necessary to control for potential differences in variances between both questionnaires that would have differently affected the scaling procedure leading to effect sizes which cannot be compared directly. Honest scores were z -standardized for both GPC questionnaire versions.

Table 15

*GPCs Matched for Social Desirability: Multivariate Regression
Results With the Condition as a Predictor of Trait Scores*

	Intercept (Honest)	Slope (Faking)
Likert		
Emotional Stability	0.00 [-0.15, 0.14]	2.67 [2.48, 2.86]
Extraversion	0.00 [-0.14, 0.14]	2.00 [1.83, 2.17]
Agreeableness	0.00 [-0.12, 0.12]	0.96 [0.78, 1.14]
Conscientiousness	0.00 [-0.15, 0.15]	3.10 [2.90, 3.30]
GPC		
Emotional Stability	-0.02 [-0.16, 0.11]	2.03 [1.86, 2.19]
Extraversion	-0.02 [-0.15, 0.11]	1.50 [1.34, 1.65]
Agreeableness	-0.03 [-0.15, 0.09]	1.22 [1.05, 1.39]
Conscientiousness	-0.03 [-0.16, 0.10]	1.94 [1.77, 2.10]

Note. Faking = factor score in faking condition; GPC = graded paired comparisons with item pairs matched for social desirability. Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Results are based on separate models for each questionnaire format (Likert and GPC).

Tables 15 shows the results for the regression of trait scores on the condition (honest vs. faking) for the Likert scales (upper rows) and the GPC questionnaire with item pairs

Table 16

GPCs Matched for Social Desirability: Multivariate Multiple Regression Results With the Interaction of Honest Scores and Format as Predictors for Faked Scores

	Intercept	Honest	Format	Honest*Format
Es	2.67 [2.51, 2.82]	0.21 [0.14, 0.28]	-0.66 [-0.85, -0.48]	-0.18 [-0.28, -0.09]
Ex	2.00 [1.89, 2.12]	0.18 [0.12, 0.25]	-0.53 [-0.68, -0.38]	-0.16 [-0.25, -0.08]
Ag	0.97 [0.82, 1.12]	0.25 [0.14, 0.37]	0.22 [0.03, 0.42]	-0.14 [-0.29, 0.01]
Co	3.11 [2.94, 3.28]	0.16 [0.08, 0.24]	-1.20 [-1.40, -0.99]	-0.15 [-0.24, -0.05]

Note. ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Honest = honest scores; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Format is dummy coded with Likert = 0 and graded paired comparisons = 1. Honest*Format is the interaction effect of these two variables.

matched for social desirability (lower rows). The results for the Likert scales are equivalent to those from the main manuscript, since the data are the same (minimal differences are presumably due to the renewed Bayesian parameter estimation). Differences of effect sizes in the GPC condition (slopes) between the GPC questionnaire with item pairs matched for social desirability (Table 15) and the full questionnaire (see corresponding table in the manuscript) are small and unsystematic. All slope CIs of both questionnaire versions overlap. Thus, the GPC questionnaire version based on item pairs matched for social desirability does not show less score inflation than the long (full) version.

Results in Table 16 remain largely unchanged. The largest changes with respect to the corresponding table for the long (full) GPC questionnaire in the main manuscript are observed for agreeableness. For this trait, the estimate for the interaction term of honest score and questionnaire format for agreeableness is now on a comparable level as for the other traits but still not significantly different from zero. Thus, the point estimate should be interpreted with caution as should any change in the estimates.

Results in Table 17 show no relevant changes with respect to our hypotheses. The point estimates for the faking intention as a predictor of faked GPC scores are somewhat weaker on a descriptive level suggesting that the intention to fake is less mirrored in the trait scores under faking conditions if the GPC items are matched for social desirability but none of these changes is significantly different from zero (credible intervals overlap). Thus, do not interpret these estimation differences.

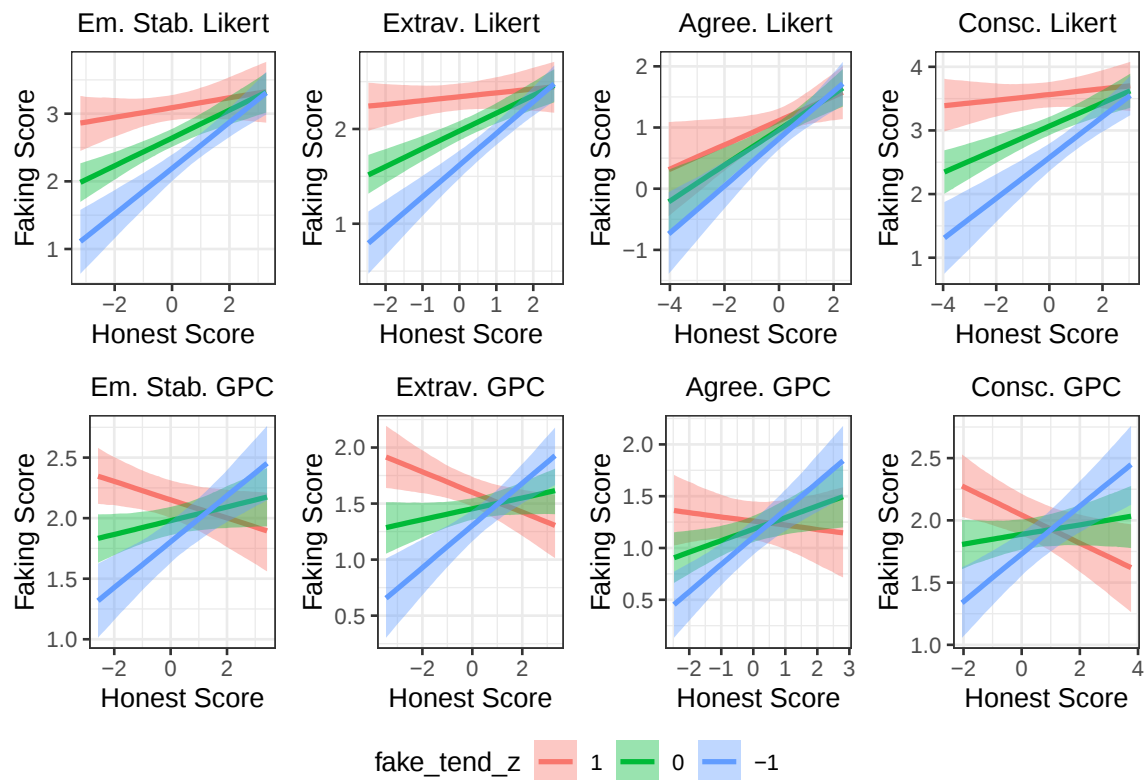
The pattern of results in Figure 7 mirror those from the long (full) GPC questionnaire in the main manuscript. One notable change is that for agreeableness measured with GPCs, honest and faking scores are now negatively correlated under high faking intention, as is the case with other traits measured with GPCs.

Table 17

GPCs Matched for Social Desirability: Multivariate Multiple Regression Results With the Interaction of Honest Scores and Individual Faking Intention as Predictors for Faked Scores

	Intercept	Honest	Faking Intention	Honest*Faking Intention
Likert				
Es	2.64 [2.51, 2.77]	0.21 [0.13, 0.29]	0.45 [0.31, 0.59]	-0.13 [-0.21, -0.05]
Ex	1.98 [1.88, 2.08]	0.19 [0.12, 0.25]	0.36 [0.26, 0.46]	-0.15 [-0.21, -0.09]
Ag	0.97 [0.83, 1.10]	0.29 [0.17, 0.41]	0.15 [0.01, 0.28]	-0.09 [-0.21, 0.02]
Co	3.07 [2.93, 3.21]	0.18 [0.11, 0.26]	0.49 [0.33, 0.64]	-0.14 [-0.21, -0.07]
GPC				
Es	1.98 [1.86, 2.09]	0.06 [0.00, 0.12]	0.17 [0.06, 0.29]	-0.13 [-0.18, -0.08]
Ex	1.45 [1.36, 1.55]	0.05 [-0.01, 0.11]	0.15 [0.06, 0.24]	-0.14 [-0.19, -0.09]
Ag	1.18 [1.06, 1.31]	0.11 [0.02, 0.20]	0.08 [-0.04, 0.20]	-0.15 [-0.23, -0.08]
Co	1.89 [1.76, 2.01]	0.04 [-0.02, 0.10]	0.16 [0.04, 0.27]	-0.15 [-0.21, -0.09]

Note. ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Honest = honest scores; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Faking intention is z transformed. Honest*Faking Intention is the interaction effect of these two variables.

**Figure 7**

GPCs Matched for Social Desirability: Interaction plots for Hypothesis 3. Em. Stab. = Emotional stability; Extrav. = Extraversion; Agree. = Agreeableness; Consc = Conscientiousness; GPC = Graded paired comparison condition.

Table 18

GPCs Matched for Social Desirability: Multivariate Multiple Regression With the Interaction of Honest Scores, Individual Faking Intention, and Questionnaire Format as Predictors for Faked Scores

	Intercept	Format	Hon Score	Intention	Hon*Form	Tend*Form	Hon*Tend	Hon*Tend*Form
	Regression coefficients							
Es	2.65	-0.67	0.21	0.44	-0.16	-0.27	-0.12	-0.01
Ex	1.98	-0.53	0.19	0.36	-0.15	-0.21	-0.13	0.00
Ag	0.97	0.21	0.26	0.14	-0.14	-0.06	-0.09	-0.07
Co	3.07	-1.19	0.18	0.49	-0.15	-0.34	-0.13	0.00
	95% CIs of regression coefficients							
Es	[2.49, 2.79]	[-0.85, -0.48]	[0.13, 0.28]	[0.29, 0.60]	[-0.25, -0.07]	[-0.45, -0.08]	[-0.19, -0.04]	[-0.10, 0.08]
Ex	[1.87, 2.09]	[-0.67, -0.39]	[0.13, 0.25]	[0.25, 0.47]	[-0.23, -0.07]	[-0.35, -0.07]	[-0.19, -0.07]	[-0.07, 0.08]
Ag	[0.81, 1.12]	[0.01, 0.41]	[0.15, 0.37]	[-0.01, 0.28]	[-0.29, 0.01]	[-0.24, 0.13]	[-0.19, 0.02]	[-0.20, 0.06]
Co	[2.91, 3.24]	[-1.38, -0.99]	[0.10, 0.26]	[0.32, 0.66]	[-0.25, -0.06]	[-0.53, -0.13]	[-0.20, -0.06]	[-0.10, 0.09]

Note. Format = questionnaire format; Hon Score = honest score; Intention = individual faking intention; ES = emotional stability; Ex = extraversion; Ag = agreeableness; Co = conscientiousness; Asteriks (*) interaction terms of the variables involved; Each column reports regression coefficients followed by the corresponding 95% credible interval [in brackets]. Faking intention is z transformed.

In Table 18, we observe no relevant changes with consequences on our hypotheses.