Weighting ratings: are people adjusting for bias in extreme reviews?

Supplemental Material

Big Five personality measure

Table S1: The 20 items of the mini-IPIP personality inventory shown to participants to elicit Big Five personality factors

Big Five Factor	Item	Polarity
Extraversion	I am the life of the party	+
Extraversion	I talk to a lot of different people at parties	+
Extraversion	I don't talk a lot	-
Extraversion	I keep in the background	-
Agreeableness	I sympathize with others' feelings	+
Agreeableness	I feel others' emotions	+
Agreeableness	I am not really interested in others	-
Agreeableness	I am not interested in other people's problems	-
Conscientiousness	I get chores done right away	+
Conscientiousness	I like order	+
Conscientiousness	I often forget to put things back in their proper place	-
Conscientiousness	I make a mess of things.	-
Neuroticism	I have frequent mood swings	+
Neuroticism	I get upset easily	+
Neuroticism	I am relaxed most of the time	-
Neuroticism	I seldom feel blue	-
Openness / Intellect	I have a vivid imagination	+
Openness / Intellect	I have difficulty understanding abstract ideas	-
Openness / Intellect	I am not interested in abstract ideas	-
Openness / Intellect	I do not have a good imagination	-

Note: a +ve (-ve) polarity means that stronger agreement with a statement corresponds to a higher (lower) level of the Big Five factor being measured.

Robustness checks for main effects in Study 1 (mixed model, fixed-effects, random effects)

Table S2: Robustness regressions testing the basic experimental treatment effect in Study 1

	Dependent variable: Reported quality of good					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Mixed	FE	FE	RE	RE
	clustered s.e.	rand slope, int	by good	by person	by good	by person
T1 dummy	2.032	2.035	1.022	2.040***	1.259	2.037***
	(1.201)	(1.194)	(1.279)	(0.461)	(1.245)	(0.450)
T2 dummy	7.929***	7.923***	5.244	7.666***	5.862**	7.750***
	(1.760)	(1.766)	(3.282)	(0.489)	(2.949)	(0.481)
Mean review score	26.71***	26.71***	26.13***	26.72***	26.18***	26.72***
	(1.179)	(1.182)	(2.372)	(0.373)	(1.245)	(0.373)
Total num reviews (N)	0.00364	0.00365	-0.0122	0.00377***	-0.0844	0.00373***
	(0.00232)	(0.00233)	(0.00868)	(0.000610)	(0.00692)	(0.000608)
Constant	-34.25***	-34.26***	-27.73**	-34.25***	-28.99***	-34.25***
	(4.716)	(4.725)	(10.286)	(1.343)	(9.516)	(1.444)
Observations	5010	5010	5010	5010	5010	5010
AIC	40444	40036	40164	38853	40222	39994
BIC	40477	40081	40190	38879	40268	40040

Notes: Regression (1) is identical to the simplest specification in Table 2 of the main paper. Subsequent regressions in this table test the same specification but using different models. The type of model is stated directly below the regression number. The mixed (multilevel) model in (2) contains a random slope and intercept parameter by good. FE = fixed-effects, RE = random-effects. Where FE and RE specifications are grouped by good, estimation is divided into 10 groups with 501 observations per group. Where FE and RE specifications are grouped by person, estimation is divided into 501 groups with 10 observations per group. Standard errors are robust to heteroskedasticity, and shown in parentheses. *** $p < .01$, ** $p < .05$, * $p < .1$. The AIC and BIC statistics for RE models must be calculated using maximum likelihood, and not feasible GLS. The RE model estimates are obtained with feasible GLS. Therefore, the AIC and BIC statistics for (5) and (6) should be considered as approximate.

Willingness to pay (WTP) regressions for Study 1

Table S3 estimates identical regression specifications to Table 2 in the main paper, again using OLS with standard errors clustered by good, but with WTP as the dependent variable. Figure S1 shows quality plotted against WTP ($r = 0.41$).

Table S3: Regressions showing the differences in WTP across treatments

	Dependent variable: Reported quality of good			
	(1)	(2)	(3)	(4)
Mean-preserving treatment (T1)	0.723	-6.452	-6.024	0.0919
	(5.756)	(3.534)	(3.516)	(11.47)
No extreme reviews treatment (T2)	18.72	-4.296	-4.329	25.11***
	(11.42)	(8.732)	(8.861)	(5.450)
Mean review score	74.51**	35.72***	35.82***	67.14**
	(26.76)	(10.69)	(10.71)	(23.53)
Total number of reviews (N)	0.0802	-0.0185	-0.0204	0.150***
	(0.0531)	(0.0419)	(0.0410)	(0.0434)
Experience good				-137.3**
				(44.61)
T1 × Experience good				8.78
				(12.80)
T2 × Experience good				6.861
				(24.07)
Dummies for each good	No	Yes	Yes	No
Demographic and Big Five controls	No	No	Yes	No
Constant	-148.0*	6.205	-31.48	-74.79
	(79.05)	(30.21)	(39.74)	(102.6)
Observations	5010	5010	5010	5010
R^2	0.147	0.519	0.524	0.334
Adjusted R^2	0.146	0.518	0.521	0.333
AIC	63318	60444	60403	62081
BIC	63351	60470	60461	62133

Notes: All regression specifications are estimated using OLS with standard errors clustered by good (i.e. 10 clusters in total). Standard errors are shown in parentheses. The set of demographic variables includes: age, sex, income, employment status, relationship status, and highest level of education. *** $p < .01$, ** $p < .05$, * $p < .1$.

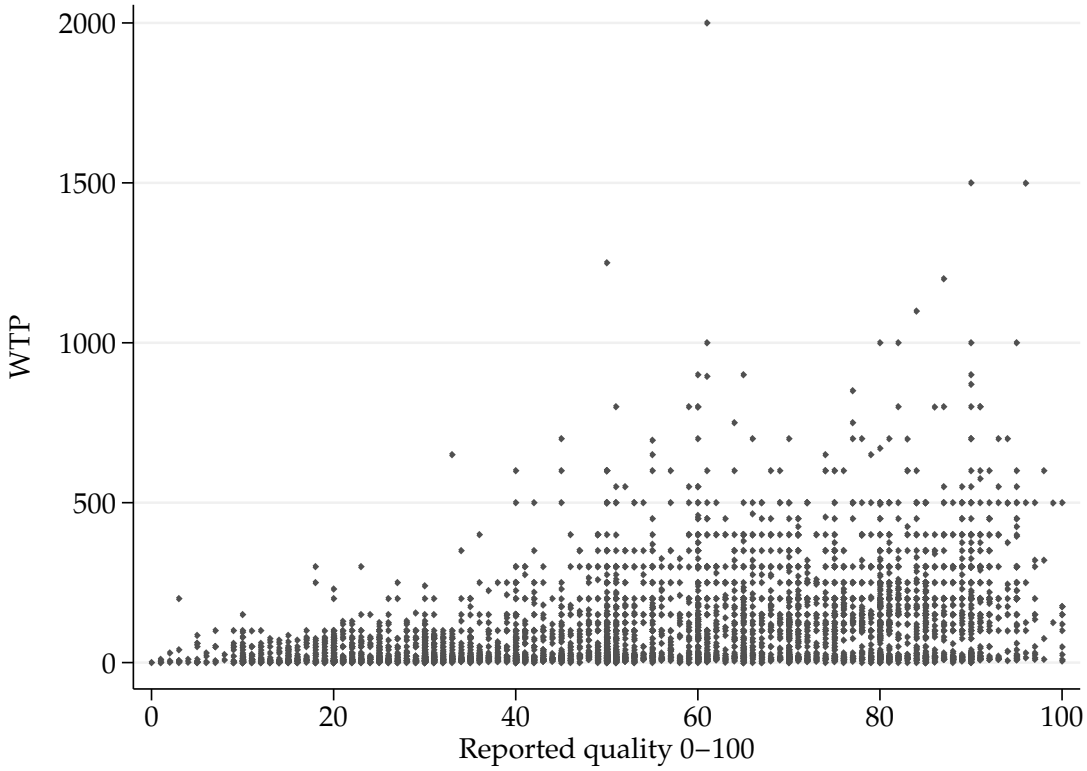


Figure S1: A plot of reported quality against WTP (in US dollars) for all observations in Study 1

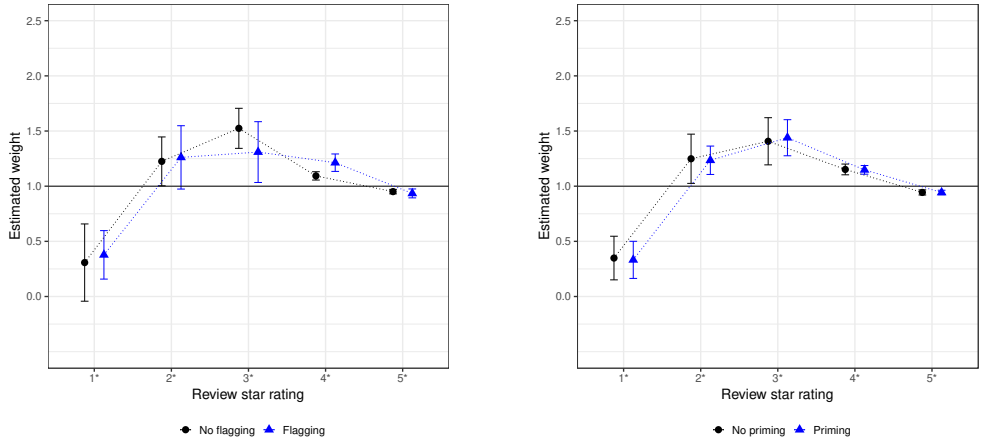
Treatment effect regressions for individual products

Table S4: Summary of average treatment effects per good from both studies

	Good number									
	1	2	3	4	5	6	7	8	9	10
Study 1										
mean-pres	0.051	-1.536	1.194	-1.809	8.609*	4.147*	-1.530	4.851*	-2.149	6.194*
s.e.	(1.625)	(1.181)	(1.529)	(1.348)	(1.401)	(1.475)	(1.044)	(1.546)	(1.425)	(1.570)
p	.975	.194	.435	.180	<.0005	.005	.143	.002	.132	<.0005
mean-pres (i.c)	-0.078	-1.283	0.625	-1.717	8.441*	4.087*	-1.978	4.428*	-2.155	5.812*
s.e.	(1.650)	(1.204)	(1.567)	(1.364)	(1.411)	(1.507)	(1.056)	(1.566)	(1.432)	(1.594)
p	.962	.287	.690	.209	<.0005	.007	.062	.005	.133	<.0005
Study 2										
priming	0.887	-0.229	0.906	-3.345*	-1.470	3.041	0.084	2.456	-1.047	-1.200
s.e.	(1.757)	(1.470)	(1.606)	(1.415)	(1.717)	(2.005)	(1.084)	(1.763)	(1.625)	(1.658)
p	.614	.876	.573	.018	.392	.130	.939	.164	.520	.470
priming (i.c)	0.659	-0.209	0.750	-3.366*	-1.737	2.623	-0.236	2.369	-1.407	-1.215
s.e.	(1.754)	(1.485)	(1.606)	(1.417)	(1.709)	(2.021)	(1.085)	(1.767)	(1.636)	(1.668)
p	.707	.888	.641	.018	.310	.195	.828	.180	.390	.467
flagging	5.520*	-2.323	2.152	-3.843*	-7.894*	3.162	-0.404	4.576*	-6.808*	4.187*
s.e.	(1.741)	(1.457)	(1.591)	(1.402)	(1.702)	(1.987)	(1.074)	(1.747)	(1.610)	(1.643)
p	.002	.111	.177	.006	<.0005	.112	.707	.009	<.0005	.011
flagging (i.c)	5.517*	-2.457	1.877	-4.282*	-8.284*	2.874	-0.648	4.555*	-7.218*	4.124*
s.e.	(1.734)	(1.468)	(1.589)	(1.401)	(1.690)	(1.999)	(1.073)	(1.747)	(1.618)	(1.649)
p	.002	.095	.238	.002	<.0005	.151	.546	.009	<.0005	.013
flag x prime	-3.040	-0.342	-1.331	5.657*	2.177	-5.225	-0.938	-3.294	1.984	0.119
s.e.	(2.490)	(2.083)	(2.275)	(2.005)	(2.434)	(2.842)	(1.536)	(2.498)	(2.303)	(2.350)
p	.223	.870	.559	.005	.371	.066	.542	.188	.389	.960
flag x prime (i.c.)	-2.678	0.057	-0.648	6.417*	2.951	-4.773	-0.339	-3.094	2.582	0.143
s.e.	(2.485)	(2.104)	(2.276)	(2.008)	(2.421)	(2.864)	(1.537)	(2.503)	(2.318)	(2.363)
p	.282	.979	.776	.002	.223	.096	.825	.217	.266	.952

Notes: Shown are estimates of the average treatment effect from OLS regressions with reported quality as the dependent variable. (i.c.) indicates estimates from regressions that control for demographics and big five personality. Standard errors are in parentheses. * indicates the treatment effect is statistically significant at the 5% level.

How weights are moderated by Study 2 treatments



(a) Flagging ($n = 3890$ for no flagging, $n = 3860$ for flagging)

(b) Priming ($n = 3960$ for no flagging, $n = 3790$ for flagging)

Figure S2: How the weighting profile is moderated by the flagging and priming treatments from Study 2