Supplementary 6 - Pilot Study

Table of Contents

1 Background

This study was the first pilot of this paradigm (before Study 1 & Study 2 that are reported in the manuscript). Its purpose was to determine whether the paradigm and the measurements are feasible, to estimate an effect size, and to find the appropriate size of violation to use in such a paradigm.
As a very exploratory study, we decided it was not yet justified to provide participants with negative evaluation on themselves. For this reason the Pilot Study included only neutral traits and feedbacks were either verifying or positively violating. To test the effect of different sizes of violation, feedback was either mildly violating or highly violating.
Besides the base paradigm, participants in the Pilot Study also answered the Rosenberg Self-Esteem Scale (Rosenberg, 1965).

2 Participant Exclusion

A total of N = 103 participants began the first session of the Pilot Study.
51 participants did not finish the study and were excluded from data analysis according to pre-registered criteria:

- Twelve participants in session 1 provided self-ratings that were unsuitable for session 2 (elaborated below).
- Fifteen participants failed the comprehension checks either at the beginning of Session 1 or at the beginning of Session 2.
- Twelve failed attention checks either on Session 1 or on Session 2.
- One did not finish Session 1.
- Six did not return to session 2 at the appropriate time or didn't finish it.
- Two participants were excluded for not understanding the instructions as evidenced by their choices.
- Three participants indicated they had ADHD and therefore did not meet inclusion criteria and were excluded from the study before they began answering.

The final analysis included N = 52 participants.

Unsuitable self-ratings-

To be eligible for the 2nd session, participants needed to provide self-rating that allow us to create feedbacks for all 3 conditions-

- Verification - a feedback similar to the self-rating (self-rating + rand(-5:5)).
- Small Positive Violation - a feedback mildly higher than the self-rating (self-rating + 15 + rand(-5:5)).
- Big Positive Violation - a feedback much higher than the self-rating (self-rating + 30 + rand(-5:5)).

To do so, an eligible participant was required to have:

- no more than 8 traits self-rated higher than 85 &
- no more than 12 traits self-rated lower than 70.

3 Stimuli - Traits and Trait Questionnaires

Traits were taken from the paper E-millim (Armony-Sivan et al., 2013), in which participants received random Hebrew words and rated them either on valence, i.e., how positive or negative the word is, or on arousal. Both scales ranged from 1-9. For the Pilot Study we chose only neutral traits, as the first stage each participant rated themselves on 20 traits in a random order (Table 1). “N valence” and “N Arousal” refer to the number of participants who rated the word on each scale. The trait questionnaires were originally written in Hebrew, for the supplementary materials Table 2 displays a rough translation to English by ChatGPT. “Original Length” refers to the number of characters (including spaces) in all 3 original Hebrew questionnaires, which we controlled for. Question order was randomized for each trait for each participant.

Table 1: Traits used in the Pilot Study as Stimuli (data from E-millim)

Trait English	Trait Hebrew	N valence	Mean valence	SD valence	N arousal	Mean arousal	SD arousal
analytical	אנליטיות	32	5.84	1.62	33	4.45	2.12
shy	ביישנות	31	4.84	1.88	33	5.24	2.53
worrisome	דאגנות	31	4.23	1.93	33	6.12	1.95
dramatic	דרמתיות	32	4.13	1.88	33	5.94	2.00
opinionated	דעתנות	27	5.56	1.95	24	6.63	2.12
indulgent	וותרנות	31	4.48	2.40	33	5.00	2.02
childish	ילדותיות	27	4.22	2.17	25	4.92	2.89
Conscientious	מצפוניות	33	5.91	2.16	33	6.09	2.22
authoritarian	סמכותיות	30	6.33	1.66	33	6.12	2.17
adaptive	סתגלנות	32	6.25	1.38	33	5.00	2.35
consistent	עקביות	27	6.67	1.34	25	5.48	2.25
stubborn	עקשנות	32	4.31	2.02	32	6.13	2.04
simple	פשטות	22	6.32	1.84	23	5.17	2.61
cool	קור-רוח	32	6.31	1.79	33	4.76	2.61
tough	קשיחות	33	4.39	2.17	32	5.84	1.72
spiritual	רוחניות	22	6.48	1.89	23	5.70	3.14
compassionate	רחמנות	33	5.82	2.23	33	5.09	2.39
rational	רציונאליות	27	6.81	1.30	25	5.24	2.37
serious	רצינות	32	6.34	1.54	33	5.52	2.18
naïve	תמימות	31	6.29	1.52	33	4.67	2.36

Table 2: Trait Questionnaires used in the Pilot Study as Stimuli

Trait English	Trait Hebrew	Question 1	Question 2	Question 3 (reversed)	Original Length
analytical	אנליטיות	I find STEM subjects easy	I find it easy to understand a logical argument in a long article	I struggle to solve arithmetic problems in my head	95
shy	ביישנות	I find it hard to speak in front of an audience	I feel uncomfortable approaching people I haven't met before	I find it easy to share personal details about my life with others	97
worrisome	דאגנות	I tend to have many worries in my daily life	I often worry about my loved ones	I find risky activities enjoyable	89
dramatic	דרמתיות	I often express emotions in an exaggerated manner	I tend to get excited by trivial events	I tend to describe events without exaggeration	97
opinionated	דעתנות	I frequently express my opinion on various topics	I tend to intervene in other people's conversations	I find it hard to express a conflicting opinion to others	99
indulgent	וותרנות	I often abandon challenging tasks	I tend to be flexible with others' decisions	I tend to insist and stand my ground	93
childish	ילדותיות	I feel bad when I make mistakes that hurt others	I find it hard to witness injustice in the world	I find it easy to stay indifferent in the face of injustice	83
Conscientious	מצפוניות	I feel bad when I make mistakes that hurt others	I find it hard to witness injustice in the world	I find it easy to stay indifferent in the face of injustice	96
authoritarian	סמכותיות	I feel comfortable giving instructions to others	I find it easy to lead and manage a group	I find it difficult to make decisions for others in a group	93

adaptive	סתגלנות	I find it easy to handle new situations	I find it easy to manage without a fixed schedule	I find it hard to cope with last-minute changes	97
consistent	עקביות	I tend to repeat similar behavior patterns	I try to maintain a consistent daily routine	I frequently change my views and opinions	98
stubborn	עקשנות	I find it hard to reach compromises with others	It's important to me to do things my own way	I can change my mind after a little persuasion	90
simple	פשטות	I tend to be satisfied with little	I tend to enjoy the simple things in life	I often buy expensive items	92
cool	קור-רוח	I find it easy to handle stressful situations	I tend to respond in a quick and matter-of-fact way	I find it hard to relax without calming exercises	94
tough	קשיחות	Insults don't affect me emotionally	I address others assertively	I'm comfortable expressing emotions to others	81
spiritual	רוחניות	I believe in a higher power	I believe I'm part of something greater	I struggle to believe in things I can't see	94
compassionate	רחמנות	I often express compassion for others' suffering	I often help people in distress	I tend to judge others harshly for their mistakes	100
rational	רציונאליות	I find it hard to perform a task that lacks logic	It's important for me to list pros and cons before making a decision	My actions are driven by emotions	95
serious	רצינות	It's important for me to keep my commitments	I find it hard to speak without prior thinking	I often laugh at myself	86
naïve	תמימות	People find it easy to deceive me	I believe the world is a good place	I find it hard to believe what others say	92

Note:
Roughly translated from Hebrew by ChatGPT. Original length refers to N symbols in Hebrew.

4 Analyses

4.1 Main Model Analysis

To model this data we used a mixed linear model with verification condition (verification, mild violation, high violation; dummy coded) predicting the desire to retake each questionnaire. The model also included a random intercept and slope for participants. As evident by the results, the size of violation the feedback caused affected the desire to retake questionnaire compared to questionnaires providing verifying feedback (Table 3 & 5). Specifically, participants forwent money to avoid retaking verifying questionnaire, did not care whether to retake or avoid mildly violating questionnaires, and forwent money to retake highly violating questionnaires (Table 4).

Table 3: Fixed Effects from the Main Model

term	β	SE	t value	df	p.value	conf.low	conf.high
Intercept	-13.777	3.936	-3.500	48.355	0.001	-21.689	-5.864
Mild Violation vs. Verification	11.751	5.121	2.295	50.591	0.026	1.468	22.033
Mild Violation vs. Verification	24.588	6.603	3.724	50.377	0.000	11.328	37.849

Table 4: Estimated Marginal Means by Verification Condition

verification_condition	emmean	SE	df	lower.CL	upper.CL
Verification	-13.777	3.939	50.582	-21.686	-5.868
Mild Violation	-2.026	4.220	49.963	-10.503	6.451
High Violation	10.812	4.939	50.410	0.893	20.730

Table 5: Effect Size (Cohen's d) for Violation vs. Verification

contrast	Cohen's d	SE	df	lower.CL	upper.CL
High Violation vs. Verification	0.732	0.197	50.410	0.337	1.128
High Violation vs. Verification	0.350	0.153	49.963	0.043	0.657

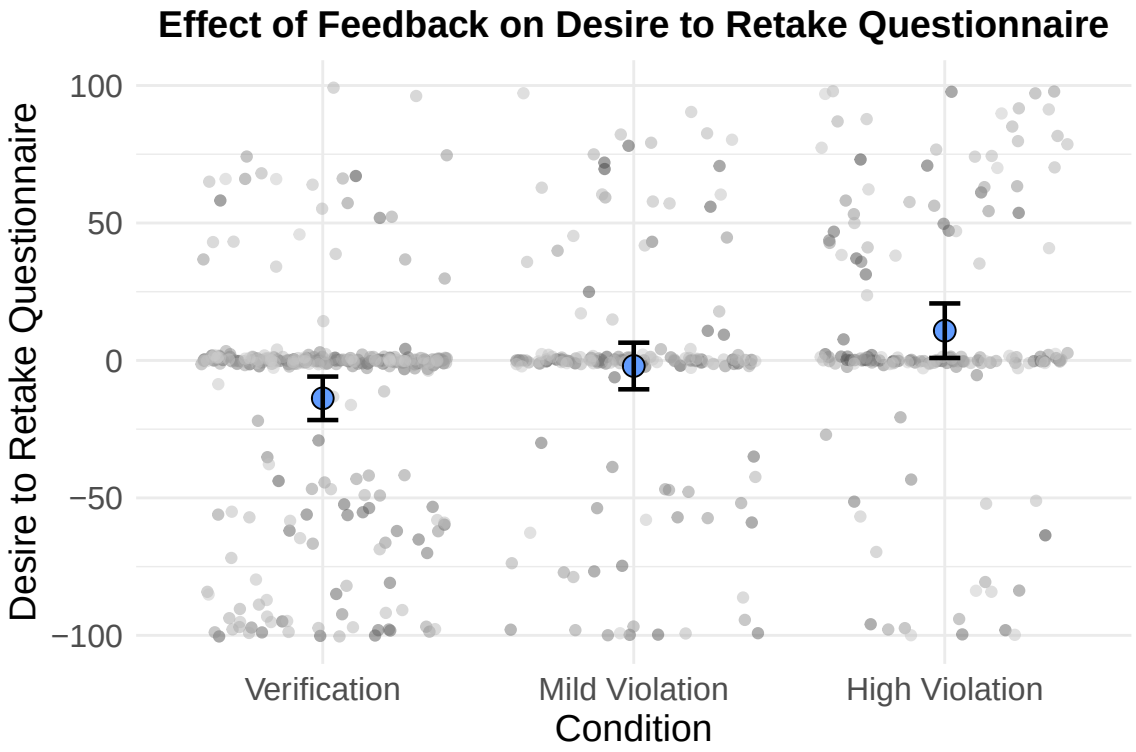

Figure 1: Gray points represent the responses at the participants' level, Blue point and error bars are means and confidence intervals derived from the Main Model

4.2 Self-Esteem Mediation

Self-esteem was measured by the Rosenberg Self-Esteem Scale. Self-esteem was coded into 3 groups as recommended in the literature (Echeburua, 1995; García at.al, 2019; Rosenberg, 1965)

As we did not have enough participants in the Low Self-Esteem group, the results should be analyzed carefully. Adding Self-Esteem to the model shrunk the effect of violation vs. verification, such that mild violation was no longer significantly different than verification, and high violation remained significantly different but with a smaller effect size. Self-esteem did not have a significant effect on the desire to retake questionnaires, nor did it have a significant interaction effect.

Table 6: RSE Group Counts

Group	Count
Low RSE (<26)	8
Medium RSE (26–29)	17
High RSE (>29)	27

Table 7: Descriptive Statistics

variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
rse.score	52	20	39	30	27	34	7	4.448	30.308	4.841	0.671	1.348

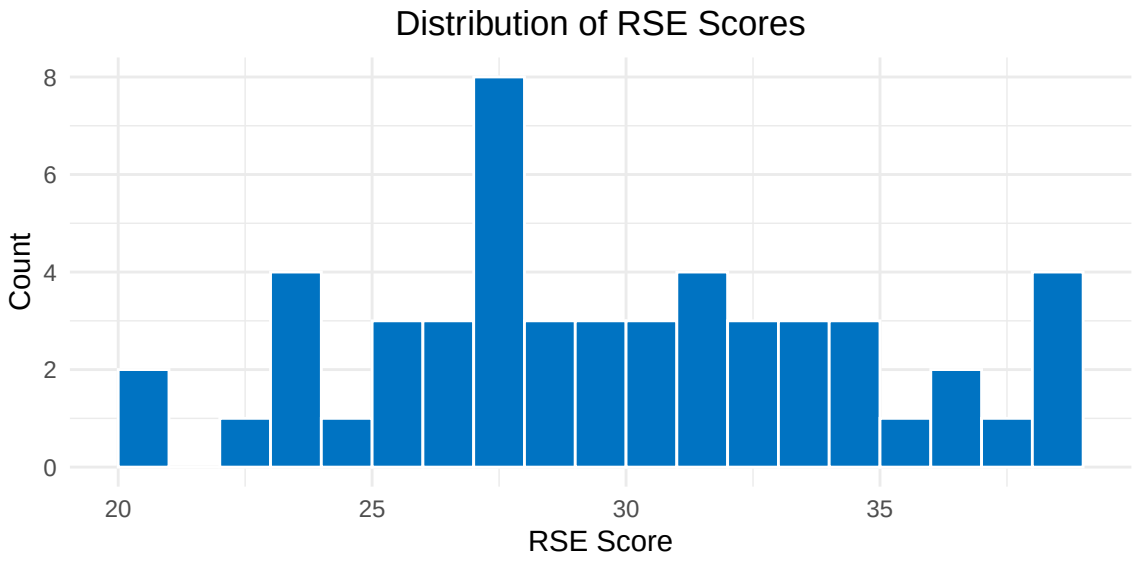

Table 8: Fixed Effects from RSE Model (Verification Condition × RSE Group)

term	β	SE	t value	df	p.value	conf.low	conf.high
Intercept	-14.433	4.428	-3.259	46.980	0.002	-23.341	-5.524
Mild Violation vs. Verification	11.445	5.891	1.943	48.071	0.058	-0.400	23.289
High Violation vs. Verification	27.103	7.509	3.609	47.777	0.001	12.003	42.203
RSE (Medium vs. Low)	-6.616	7.322	-0.904	48.279	0.371	-21.336	8.104
RSE (High vs. Medium)	9.591	5.887	1.629	45.300	0.110	-2.264	21.445
Mild Violation × RSE (Med vs. Low)	-3.113	9.648	-0.323	47.535	0.748	-22.517	16.291
High Violation × RSE (Med vs. Low)	9.343	12.312	0.759	47.454	0.452	-15.420	34.105
Mild Violation × RSE (High vs. Med)	2.690	7.934	0.339	48.838	0.736	-13.256	18.636
High Violation × RSE (High vs. Med)	-5.031	10.097	-0.498	48.243	0.621	-25.330	15.268

Table 9: Estimated Marginal Means: Condition × RSE Group

verification_condition	rse_group	emmean	SE	df	lower.CL	upper.CL
Verification	Low	-21.049	10.106	52.298	-41.324	-0.773
Mild Violation	Low	-12.717	10.395	43.605	-33.672	8.238
High Violation	Low	15.397	12.356	44.981	-9.489	40.284
Verification	Medium	-4.842	6.722	46.141	-18.373	8.688
Mild Violation	Medium	9.292	7.487	51.902	-5.732	24.317
High Violation	Medium	17.229	8.785	51.401	-0.404	34.863
Verification	High	-17.408	5.422	49.192	-28.302	-6.513
Mild Violation	High	-5.540	5.784	46.961	-17.176	6.097
High Violation	High	5.384	6.836	47.676	-8.364	19.131

Table 10: Estimated Marginal Means: RSE Group

rse_group	emmean	SE	df	lower.CL	upper.CL
Low	-6.123	7.347	47.106	-20.901	8.656
Medium	7.227	5.127	50.321	-3.070	17.523
High	-5.854	4.030	48.476	-13.955	2.246

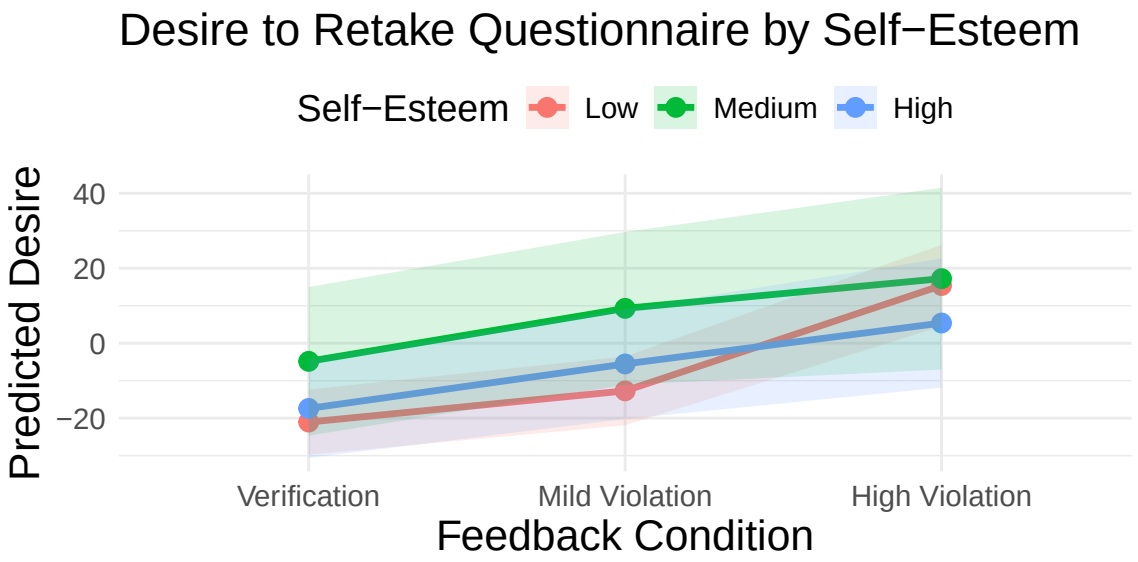

Figure 2: Interaction between Feedback, RSE on desire to retake questionnaire. The plot displays predicted data according to modelbased::estimate_relation, and not original data.

4.3 Trait Centrality Model

For each trait, participants rated how central that trait is to them. Centrality was significantly correlated with trait self-rating

Pearson correlation: r(622) = 0.43, t = 12.03, p < .001.

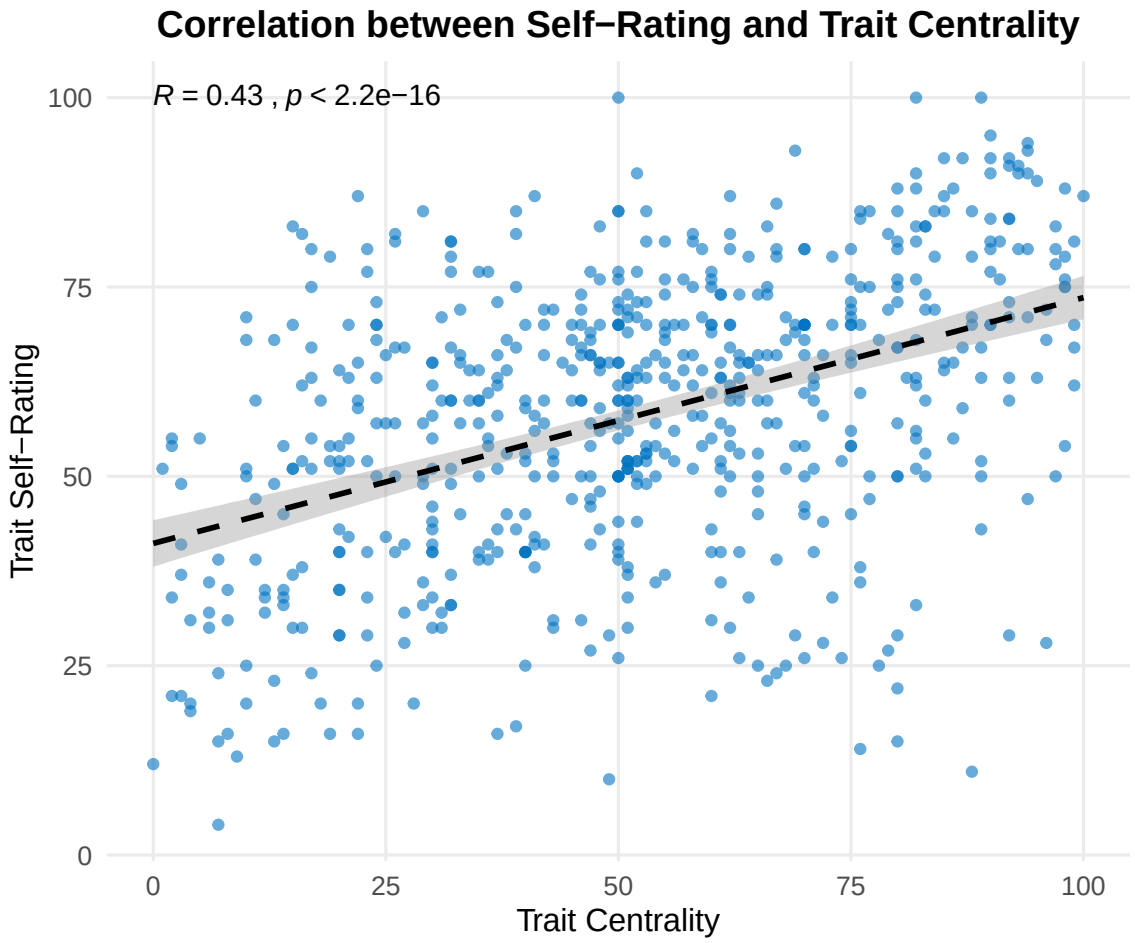

Figure 3: Correlation between self-rating and trait centrality

To help the interpretation of the centrality model, we mean-person centered centrality.

Adding trait centrality (centered) to the main model did not impact the significance of the effects of violations versus verification, but reduced their effect sizes. Neither the trait centrality nor its interaction with the feedback had a significant or consistent (across studies) effect on the desire to retake questionnaires.

Table 11: Fixed Effects from Centrality Model

term	β	SE	t value	df	p.value	conf.low	conf.high
Intercept	-13.646	3.939	-3.465	48.689	0.001	-21.562	-5.730
verification_condition2	11.593	5.145	2.253	50.677	0.029	1.261	21.924
verification_condition3	24.278	6.645	3.653	51.563	0.001	10.941	37.615
Centrality	-0.044	0.090	-0.491	496.436	0.624	-0.220	0.132
verification_condition2:centrality	0.097	0.159	0.606	517.453	0.545	-0.216	0.410
verification_condition3:centrality	0.014	0.152	0.093	522.439	0.926	-0.285	0.313

Table 12: Effect of Centrality on Desire to Retake Questionnaire by Feedback Condition

verification_condition	centrality.trend	SE	df	lower.CL	upper.CL
Verification	-0.044	0.090	497.136	-0.220	0.132
Mild Violation	0.053	0.132	473.031	-0.207	0.312
High Violation	-0.030	0.125	518.042	-0.274	0.215

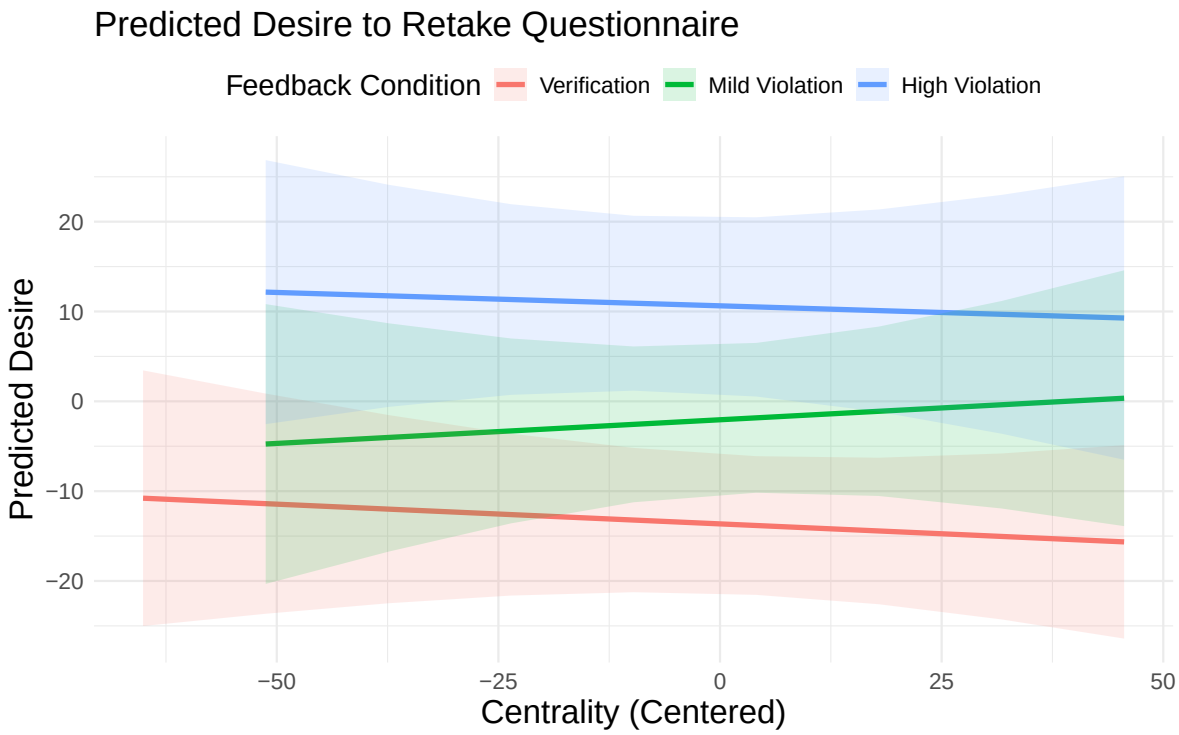

Figure 4: Interaction between Feedback and Trait Centrality on desire to retake questionnaire. The plot displays predicted data according to modelbased::estimate_relation, not original data.