Supplementary Table 1 Reaction times between the MDD group and healthy controls under different control conditions

Effect	*df*	*F* value	*P* value	ηp^2
Group (MDD vs HCs)	1, 54	0.702	0.406	0.013
Stereotype compatibility	2, 108	65.363	< 0.001[b]	0.548
Gender bias type	1, 54	6.916	0.011[a]	0.114
Group × Stereotype compatibility	2, 108	0.753	0.473	0.014
Group × Gender bias type	1, 54	2.303	0.105	0.041
Stereotype compatibility × Gender bias type	2, 108	4.105	0.048[a]	0.071
Group × Stereotype compatibility × Gender bias type	2, 108	1.477	0.134	0.016

[a]$P < 0.05$.

[b]$P < 0.01$.

Reaction time data were analyzed using a 2 (group: MDD vs healthy controls) × 3 (stereotype compatibility: congruent, incongruent, neutral) × 2 (gender bias type: male, female) mixed-design repeated-measures ANOVA (n = 56). Post-hoc comparisons were conducted using Bonferroni correction. MDD = major depressive disorder; HCs = healthy controls.

Supplementary Table 2 Comparison of N170 amplitude and latency in MDD group and healthy controls

	Effect	*df*	*F* value	*P* value	ηp^2
Amplitude	Group (MDD vs HCs)	1, 54	0.121	0.730	0.002
	Stereotype compatibility	2, 108	8.302	0.001[b]	0.129
	Gender bias type	1, 54	7.401	0.009[b]	0.121
	Group × Stereotype compatibility	2, 108	1.078	0.344	0.020
	Group × Gender bias type	1, 54	2.109	0.152	0.038
	Stereotype compatibility × Gender bias type	2, 108	7.113	0.001[b]	0.116
	Group × Stereotype compatibility × Gender bias type	2, 108	0.674	0.512	0.012
Latency	Group (MDD vs HCs)	1, 54	1.641	0.206	0.029
	Stereotype compatibility	2, 108	0.068	0.934	0.001
	Gender bias type	1, 54	5.436	0.023[a]	0.091
	Group × Stereotype compatibility	2, 108	0.827	0.440	0.015
	Group × Gender bias type	1, 54	2.375	0.129	0.042
	Stereotype compatibility × Gender bias type	2, 108	2.819	0.640	0.050
	Group × Stereotype compatibility × Gender bias type	2, 108	0.214	0.808	0.004

[a]$P < 0.05$.

[b]$P < 0.01$.

N170 amplitude and latency were analyzed using a 2 (group: MDD vs healthy controls) × 3 (stereotype compatibility: congruent, incongruent, neutral) × 2 (gender bias type: male, female) mixed-design repeated-measures ANOVA (n = 56). Post-hoc comparisons were conducted using Bonferroni correction. MDD = major depressive disorder; HCs = healthy controls.

Supplementary Table 3 Comparison of N400 amplitude and latency in MDD group and healthy controls

	Effect	*df*	*F* value	*P* value	ηp^2
Amplitude	Group (MDD vs HCs)	1, 54	8.937	0.004[b]	0.142
	Stereotype compatibility	2, 108	10.624	0.002[b]	0.164
	Gender bias type	1, 54	4.413	0.037[a]	0.068
	Group × Stereotype compatibility	2, 108	1.078	0.344	0.020
	Group × Gender bias type	1, 54	6.330	0.015[a]	0.105
	Stereotype compatibility × Gender bias type	2, 108	7.899	0.024[a]	0.051
	Group × Stereotype compatibility × Gender bias type	2, 108	8.323	0.006[b]	0.134
Latency	Group (MDD vs HCs)	1, 54	0.207	0.651	0.004
	Stereotype compatibility	2, 108	6.913	0.027[a]	0.084
	Gender bias type	1, 54	0.372	0.544	0.007
	Group × Stereotype compatibility	2, 108	0.928	0.398	0.017
	Group × Gender bias type	1, 54	1.613	0.210	0.029
	Stereotype compatibility × Gender bias type	2, 108	0.606	0.547	0.011
	Group × Stereotype compatibility × Gender bias type	2, 108	0.270	0.765	0.005

[a]$P < 0.05$.

[b]$P < 0.01$.

N400 amplitude and latency were analyzed using a 2 (group: MDD vs healthy controls) × 3 (stereotype compatibility: congruent, incongruent, neutral) × 2 (gender bias type: male, female) mixed-design repeated-measures ANOVA (n = 56). Post-hoc comparisons were conducted using Bonferroni correction. MDD = major depressive disorder; HCs = healthy controls.

Supplementary Table 4 ANCOVA results of N400 amplitude and latency after adjusting fluoxetine equivalent dose

	F value	*P* value	ηp^2
Amplitude			
Group (MDD vs. HCs)	8.721	0.005[b]	0.141
Stereotype compatibility	10.387	0.003[b]	0.162
Gender bias type	4.259	0.044[a]	0.074
Group × Stereotype compatibility	1.792	0.167	0.035
Group × Gender bias type	6.183	0.016[a]	0.104
Stereotype compatibility × Gender bias type	7.754	0.026[a]	0.049
Group × Stereotype compatibility × Gender bias type	8.126	0.007[b]	1.133
Latency			
Group (MDD vs. HCs)	0.198	0.665	0.004
Stereotype compatibility	6.782	0.029[a]	0.082
Gender bias type	0.369	0.551	0.007
Group × Stereotype compatibility	0.915	0.405	0.016
Group × Gender bias type	1.581	0.214	0.029
Stereotype compatibility × Gender bias type	0.596	0.557	0.011
Group × Stereotype compatibility × Gender bias type	0.269	0.764	0.005

[a]$P < 0.05$.

[b]$P < 0.01$.

Analysis of covariance (ANCOVA) was conducted to examine group differences in N400 while controlling for fluoxetine-equivalent dose (mg/day), which was included as a covariate. MDD = major depressive disorder; HCs = healthy controls.

Supplementary Table 5 Correlations between clinical/medication variables and ERP amplitude in the MDD group

Analysis family	Variable	ERP amplitude	Condition	*r* value	*P* value	FDR-adjusted *P* value
Disease severity	Disease duration	N400	Female stereotype–male face	0.560	0.003	0.036[a]
	HAMA	N170	Neutral–female face	-0.440	0.029	0.177
Medication dose	Fluoxetine equivalent dose	N400	Male stereotype–male face	0.368	0.021	0.126
	Fluoxetine equivalent dose	N170	Neutral–female face	0.403	0.032	0.192

[a]$P < 0.05$.

Spearman correlation analyses were conducted in the MDD group. Disease severity-related analyses included disease duration, HDRS-24, and HAMA scores. Medication-related analyses included the fluoxetine equivalent dose. Multiple comparisons were controlled using the false discovery rate (FDR) procedure. Only correlations with uncorrected $P < 0.05$ are reported for clarity. Both uncorrected and FDR-adjusted *P* values are presented. MDD = major depressive disorder; HAMA = Hamilton Anxiety Rating Scale; ERP = event-related potential.