within\_day\_reliability

Giacomo Bignardi

2023-10-13

## S7. Within-day test-retest reliabilities for taste-typicality and evaluation-bias.

Since some degree of unknown unreliability of taste-typicality and evaluation-bias can upwardly bias, and hence confound, estimates of unique environmental effects, we quantified reliabilities () of such metrics. First, we computed taste-typicality and evaluation-bias separately and only for the repeated images (15 for abstract images, 15 for scenes, and 60 for faces). In the discovery sample, Pearson correlations for the taste-typicality and evaluation-bias within individuals, computed on only one pair per twin, were *r* = 0.83, 95% CI [0.8, 0.85], *r* = 0.8, 95% CI [0.77, 0.82], and *r* = 0.78, 95% CI [0.75, 0.8] taste-typicality, and *r* = 0.92, 95% CI [0.91, 0.93], *r* = 0.93, 95% CI [0.92, 0.94], and *r* = 0.97, 95% CI [0.97, 0.97] for evaluation bias, for abstract images, scenes, and faces, respectively. These results were similar to the validation sample, with Pearson correlations of *r* = 0.77, 95% CI [0.74, 0.8] and *r* = 0.84, 95% CI [0.81, 0.86], and *r* = 0.93, 95% CI [0.92, 0.94] and *r* = 0.96, 95% CI [0.95, 0.96] for evaluation-bias (scenes and faces, respectively).

To get a more precise and relevant estimate of the reliability of the metrics computed on the averages of the repeated metrics, as we have done in our main analyses, we also used Intra Class Coefficient (*ICC*) of taste-typicality and evaluation-bias to estimate . In the discovery sample, taste-typicality were all good; *ICC(2,k)* =.90, *ICC(2,k)* =.89, and *ICC(2,k)* =. 87, while evaluation bias were all excellent, *ICC(2,k)* = 0.95, *ICC(2,k)* = 0.96, and *ICC(2,k)* = 0.98. Results were nearly identical in the validation sample: with *ICC(2,k)* =.87 and *ICC(2,k)* =.91, for taste-typicality, and *ICC(2,k)* = 0.96 and *ICC(2,k)* = 0.98, evaluation-bias (scenes and faces, respectively). These results indicate good reliability of taste-typicality and evaluation-bias.