

Supplementary Methods

Tests of data quality

Overview

We analysed the risk of individuals submitting low quality data in three ways. First we included a simple dominance test to check for irrational responses. Second, we identified participants who exhibited a systematic choice pattern, e.g. systematically selecting the 1st option as “most preferred”. Third, we analysed the response time for each choice task, allowing identification of participants whose responses to the choices were either much quicker (possibly representing “clicking through” choices rather than stopping to consider) or slower (possibly indicating either difficulty comprehending the choices or distraction by other activities).

Dominance

In addition to the 12 choice tasks, each DCE included two additional tasks: a “warm-up” choice task and a “dominance” test. The former was used to familiarise participants with the format of the choice questions and act as a transition between the instructions and first live choice task. The dominance task was the last task faced by the participants and included three options: A) all attributes set at their “best” levels (high level of personalisation with lowest level of cost); B) all attributes set at their “worst” levels (neutral personalisation with highest level of cost); and C) an intermediate option. Participants were expected to choose rationally, thus alternative A as “most preferred” and alternative B as “least preferred”. An example choice task is shown in Web Appendix A. Dominance: A respondent was considered as failing dominance test when she/he selected an irrational, “wrong”, alternative for both best and worst choice.

Systematic choice bias

We measured serial non participation (SNP), the systematic choice based on some criterion other than the content of the attributes in relation to choice order. A respondent was considered as a serial non participant when at least 75% of his/her choices were on the same choice position (e.g. first presented, last presented) for at least one type of decision (either Best or Worst)

Response time

For each choice task we computed the 1st quintile (20%), the median (50%) and the 4th quintile (80%) of response time (RT). We then used a bootstrapping procedure with 1000 replications to obtain the 95% confidence interval (CI) around the {20%; 50%; 80%} measures. A response time was classified as a fast outlier when its duration was shorter than the lower bound of the 95% confidence intervals for the 1st RT quintile or classified as a slow outlier when its duration was longer than the upper bound of the 95% confidence intervals of the 4th RT quintile. A respondent was considered as an outlier when either $\geq 50\%$ of response times to decision tasks (i.e. both best and worst decisions for each choice task) were fast outliers, or $\geq 50\%$ of response times to decision tasks were slow outliers.