*What's the main question being asked in this study?*

When facing difficult decisions with many options and limited time, people will generate a small subset of those options – a consideration set – to actually evaluate, and they will be more likely to include options that had a high subjective value in the past. We test this in an experiment with two stages. In Stage 1, people are trained to associate a series of English words with different randomly assigned values. In Stage 2, people use these words as potential answers to a difficult decision. We analyze their answers to reveal a unique pattern predicted by our model of consideration sets.

*Describe the key dependent variables specifying how they will be measured.*

There is one key dependent variable: which word people choose in Stage 2. First, we will describe the decision that needs to be made: "Give us a word from Stage 1 whose third letter is late in the alphabet. You'll win points based on the position of the word in the alphabet (i.e. A = 1, Z = 26)." We will then give an example, two comprehension checks, and then a textbox to submit a response. (Participants will also get a “scratchpad” textbox to help them think.) To parse participants' answers, we will compare their response to the list of Stage 1 words using the Optimal String Alignment method in the "amatch" function of R package "stringdist" (with a maximum distance of 2). If their response fails to match a Stage 1 word, we will try the same parsing procedure on the scratchpad. If both responses fail to match a Stage 1 word, the trial will be coded as NA. The decision will have a 25-second time limit.

After Stage 2, we will also give participants a free recall test, where they are asked to recall as many of the Stage 1 words as they can.

*How many and which conditions will participants be assigned to?*

There are two between-subject conditions. In the “small set” condition, there are only four words in Stage 1; their Stage 1 and Stage 2 values are (13, 3, 8, 7) and (11, 11, 21, 2). We label these words “high S1”, “low S1”, “high S2”, and “distractor”, respectively. In the “large set” condition, there are twelve words in Stage 1; their Stage 1 and Stage 2 values are (13,3,8,1,5,6,7,9,10,11,15) and (11,11,21,2,2,2,2,2,2,2,2), respectively. (This is the same set, with eight “distractor” words instead of one.) The specific words are randomly drawn for each participant from a large pool that share the relevant Stage 2 value.

*Specify exactly which analyses you will conduct to examine the main question.*

(A) We predict that people will be less likely to choose the “high S2” word in the “large set” condition compared to the “small set” condition. This will be tested by fitting a logistic regression model, where the dependent variable is a dichotomized version of people’s choices – either “high S2” or “not high S2” – and the regressor is condition (“large” vs “small” set). We will compute a one-tailed p value.

(B) We predict that people will be more likely to choose the “high S1” word in the “large set” condition compared to the “small set” condition. This will be tested in an analogous manner to (A).

(C) We predict that the increase in choice probability for the “high S1” word in the “large” versus “small” conditions will be greater than the increase in probability for the “low S1” word. More precisely, we predict that the difference in log odds for the “high S1” word in the “large” versus “small” conditions will be greater than the difference in log odds for the “low S1” word in the two conditions. To test this, we will run a permutation test. We will shuffle the condition labels across subjects 10,000 times, each time computing the key test statistic: the difference in log-odds-differences (i.e. (log-odds(high S1, large) – log-odds(high S1, small)) – (log-odds(low S1, large) – log-odds(low S1, small))). Then, we will derive a one-tailed p value by computing the percentile of the actual test statistic (i.e. the difference in log-odds-differences from the actual data sample) in this null distribution.

*How many observations will be collected?*

N = 500 (before exclusion). Bootstrap simulations from pilot data indicated that this sample size would give us over 80% power.

*Anything else you would like to preregister?*

We will exclude participants for whom any of the following is true: They don't complete the study, they choose the better alternative in Stage 1 training on less than 70% of trials, they fail to give a word within the time limit in Stage 2, they fail the Stage 2 comprehension checks, they fail to recall the three non-distractor words (“high S1”, “low S1”, and “high S2”) in the final free recall test, or they write things down physically during the experiment (as measured by a probe at the end).

We will conduct a secondary analysis to test whether Stage 1 value influences the order of recall in the free recall test. Specifically, we predict that people will, on average, recall higher Stage 1 value words before lower Stage 1 value words. To test this, we will estimate a linear mixed effects model, regressing the order of recall on Stage 1 value (with random slopes and intercepts for each subject); we will compute a one-tailed p value.