Pilot Study: Building and validating the open-ended alignment stimuli.

Methods

Participants

We tested participants in a simple verbal task, in which they were asked to name items in specific categories. Pilot study asked participants to give answers that were obvious to them personally. In Study 1 participants were explicitly instructed to try to align with a group. Participants were not able to participate in both studies.

Participants in both studies were split into two age groups, older (aged 55-65) and younger (25-35). These age groups were chosen to roughly correspond to the social categories of 'millennials' and 'boomers' as we expected there to be large cultural differences between these age groups. The age groups do not map directly onto the social categories as they were reduced slightly to ensure both age ranges were the same size and contained adults who were of working age.

All participants were recruited online via Prolific and paid at an average rate of £7.57 per hour (specific hourly rate depended on how quickly they completed the study). All participants spoke English as a first language, self-reported their nationality as British, and were white. We limited our sample to white-British participants to avoid confounding age with nationality and race. At the time of sampling there were large differences in the demographic profile of participants on Prolific, where younger participants were diverse in nationality and race, whereas older participants were almost exclusively white-British. We will return to this decision and the limitations it places on interpretation in our discussion.

After completing the verbal task, all participants in both studies also took part in a visual abstract task. This condition was included as an opportunity for piloting and exploration of a new class of stimuli. We did not intend to compare or combine analysis of data from the visual and verbal tasks (see pre-registration) and because participants always completed the verbal tasks first their performance on the verbal tasks cannot have been influenced by performance on the visual tasks. The detailed method for the visual task can be obtained in the pre-registrations. The results for the visual task will not be described.

Procedure

In order to select stimuli for alignment that had differing age-related context, we first asked 40 older and 40 younger participants to suggest answers for 62 questions such as "Name a book" or "Name something that might be found on the beach" (a full list of questions is given in the SI and code to reproduce the questionnaire can be found via the OSF). In this first study we did not ask participants to align, but asked them to give answers that were obvious to themselves - primary salience. We specifically asked for primary salience in this first study in order to identify items where younger and older adults had different primary intuitions.

All participants saw all questions in a fixed randomised order. For every item, participants

were told “Type a suggestion for the category below that you think is the most obvious example of something in this category”. Items were presented one at a time, and answers could be freely typed into a text-box. It was possible to submit any answer, but not possible to leave the box blank.

After item 20 and item 40 a multiple-choice attention-check question was presented, and participants were given the option to have a self-paced break.

Results

Alignment

Alignment scores were calculated for each participant for each question by counting the total number of matching answers (i.e. answers given by other participants that are the same as their own answer) given to that question by the participant’s own age group. The number of matches is then divided by the total number of other participants in that age and condition. For example, 10 younger participants suggested ‘Harry Potter’ for the item ‘Name a book’. A younger participant who gave *Harry Potter* as their answer would match with 9 group members, so their alignment score for this item would be 9/(40-1) = 0.23.

To calculate if alignment was above chance, a baseline level of alignment was approximated for each item (within each age group) by taking 1/the total number of unique suggestions. For example, there were 24 unique suggestions made by older people for the item ‘Name a book’ so the chance level for ‘Name a book’ for older participants was 1/24=0.04.

Alignment scores for both age groups are given in table 1. Although they were not asked to align, mean alignment scores were significantly above a mean chance level for both age groups ($p<0.001$ for both) when taking all questions together. However, there was a huge amount of inter-item variability. If running the comparisons for each individual item separately there were 17 items where alignment was not above chance levels for one or both age groups ($\alpha=0.05$, not corrected for multiple comparisons, see SI for list of items).

Table 1: Mean Alignment scores for older and younger participants in study 1. These scores are a mean across all items

Age Group	**Mean Alignment Score**
Older	0.23
Younger	0.23

Identifying high- and low-context items

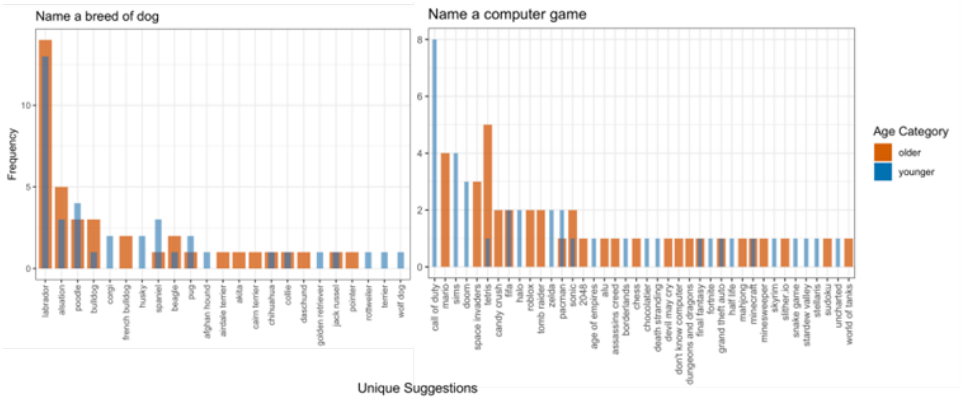

Figure 1: Examples of the distribution of answers in a low-context item (left) and a high-context item (right)

To select items as high- or low-context we first removed items where alignment on a single answer was so high that there was no variation in the data. Alignment was deemed as too high for a particular item if the mean alignment score for that item was over 0.5 and the majority answer for that question was picked by over 75% participants. From the remaining items, we then identified low-context items as items where the most popular suggestion was the same in both age groups, *and* the difference between the proportion of participants who selected that answer was less than 0.2, *and* the percentage difference between proportion of participants who selected that answer was less than 50%. This was to differentiate items that had the same majority answer, but for one age group this was very clearly favoured over other items and for the other age group there was a broad spread across other suggestions.

This selection procedure resulted in the removal of four items for having extremely high levels of alignment. Thirty-three items were classified as having low age-related context. The 25 remaining items were classified as high-context. In order to have an equal number of items at each context level, only the 25 items that varied the least between ages were retained in the low-context set.

Figure 1 shows an example of the response distributions for a high- and low-context item