This experiment was conducted in the FEELE lab at the University of Exeter using pen and paper. Before each session started, we pre-configured the 32 desks in the FEELE laboratory with the experimental materials. Every desk had the same set of items, which consisted of a large clear seal-able document envelope, a small white envelope containing twenty 20 pence coins, three $1 coins and two white envelopes, one marked with a letter A, the other marked with a letter B. An image of the configuration of one of the desks is given in Figure 7 in Appendix B. Each desk was clearly labelled with a desk number from 1 to 32.

Upon entering the lab, each subject was given a blank participant payment receipt form, drew a numbered token at random and sat at the desk labelled with the token number they drew. Subjects were asked to fill in their name on the payment receipt and sign it before the experiment began. Subjects were then asked to take the booklet out of envelope A and the instructions from the first page of the booklet were read aloud.

The instructions detailed that subjects should not write their name or any other identifying information on the lab materials other than in the payment form that they had already signed to ensure the anonymity of responses. Subjects were informed that the experimenters would not examine the payment receipts; these were to be processed only by the university's finance team. The instructions stated that the experiment would have two parts, the first of which was a series of questions in a booklet attached to the instructions sheet subjects were reading. Subjects had ten minutes to complete the booklet, which contained questionnaire on personal characteristics, a risk attitude elicitation task, an intertemporal investment task and a cognitive reflection test.

At the end of the ten minutes, subjects were asked to put the booklet for the first part of the experiment to one side and open envelope B. Again the instructions from the first page of the booklet were read out loud. The instructions for the second part of the experiment detailed a `numeracy test'. The test consisted of a series of calculations of the form x + y 􀀀 z =?. Subjects had 30 seconds to complete as many calculations as possible out of a total of twenty; each solution was worth $0.20. At the end of the 30 seconds, subjects had to check how many calculations they had answered and pay themselves accordingly from the envelope of $0.20 coins on the desk. Subjects then had to palce the two booklets for the experimental tasks, as well as the envelope containing any remaining coins into the clear plastic bag and seal it. We asked subjects to complete the payment form with the total amount for the experimental session, but to keep that separate from the other materials.

Once subjects completed both tasks they were free to leave. We asked subjects to place the clear plastic bag with the experimental materials in one box, and the payment receipt face down in another box on their way out of the lab. Both boxes were located by the exit door to the lab. Cheating was possible since subjects could keep any amount they wished before returning the experimental materials to the boxes on their way out of the lab. We could detect cheating at the individual level because both the experimental materials and the envelope used to return any cash had a (visible) ID code. However, it was impossible for the experimenters to match ID numbers to individual participants: the payment receipts purposely did not have an ID number, and the fact that around 30 subjects left at the same time meant it would be extremely unlikely for both sets of documents to be in the same order.

We ran the first three sessions under a baseline condition, which we will henceforth denote as CONTROL. The distribution of reports from those three sessions (measured in the manner described in the next paragraph) had two modes, with 18% of subjects reporting a value of 10 and another 18% reporting a value of 20. These modal responses from CONTROL were selected for use in the two main treatments. We conducted three further sessions for each of two additional treatments (leading to a total of nine sessions) using one of the two modes from the initial control treatment to induce a reference point.

In the first treatment, LOW, an additional line was added to the instructions that stated:

In previous sessions using this test one of the two most commonly reported values was 10 questions.

In a second treatment, HIGH, the additional line stated instead:

In previous sessions using this test one of the two most commonly reported values was 20 questions.

After each session ended, we opened the plastic envelopes returned by each of the subjects and inspected the contents. We counted the number of coins each subject returned and recorded and calculated the score reported by a subject as 20 minus the number of coins returned. We marked the numeracy test and recorded the number of questions attempted, as well as the number of correct answers by a subject.