# Experiment 1A

We devised both a laboratory- and an online version of the present experiment. Small differences exist between the two versions, and this allowed to merge their data. Such differences are detailed in the following sections.

## Materials and Methods

### Participants

Thirty-three students (16 females) from RWTH University took part in the laboratory version of the experiment. Other 13 students took part in the online version, together with 83 participants recruited via Prolific (<https://www.prolific.co/>). Psychology students were recruited via email and word of mouth and were compensated with partial course credits. Participants in Prolific were pre-screened to be between 18 and 35 years old and were rewarded with ₤2.75 (₤6.6/h).

Overall, there were 74 females (2 did not declare), mean age was 23.8 (± 3.6) years, and 14 participants were left-handed (one did not declare)[[1]](#footnote-1).

Participants with average error rate above 20% were removed from further analyses (1 lab participant and 4 online participants). Thus, our final sample included 124 participants with an average error rate equal to 7% (±4%).

### Stimuli, tasks, and responses

The lab experiment was run using Psychopy v1.84. Instructions and stimuli were presented in black over a grey background. Participants looked at the screen from approximately 50 cm distance. Target stimuli were digits from 1 to 9, excluding five. They were approximately 8 mm high and 4 mm wide and were presented in the middle of the computer screen (19 inches, screen resolution: 1280, 1024 pixel). The cue was a filled rectangle, 3.5 cm wide and 1.47 cm high, centred at 7.5 cm above the centre of the screen. There was a black empty frame surrounding both the cue and the stimulus, that appeared and disappeared together with the cue in each trial. The frame was a 24 cm square, centrally presented. Cue and frame were such that the number of pixels occupied by the cue was equal to occupied by the frame. Participants were required to perform a magnitude task, indicating whether the stimulus was less or greater than 5, or a parity task, indicating whether the stimulus was odd or even. The task to perform in each trial was signalled by cue orientation, that could be either horizontal or vertical. Cue-task mapping was counterbalanced across participants. Participants responded using A and L keys on a German QWERTZ keyboard, so that response-key mapping overlapped across the two tasks. Response-key mapping was counterbalanced across participants.

The online experiment was hosted and run on Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Task and stimuli were the same as for the lab experiment, but the cue and the frame were smaller (cue was 2 x 1 cm and frame side was 11.5 cm) to ensure that these were visible on laptops with average to small screen size. The digits had the same size as for the lab experiment. The instructions and the digits were black, as in the lab version, however the background was white, and not grey, throughout the whole online experiment.

Our operationalization of a task-irrelevant contextual feature consisted in assigning a colour to the cue, so that the filled rectangle could be red or blue. Since the task to be performed was signalled by the orientation of the rectangle (horizontal versus vertical), its colour didn’t provide any information functional for task recognition and response selection; in facts, it was a task-irrelevant feature enriching the context in which participants completed our task switching paradigm. A further manipulation was applied to the task-irrelevant contextual feature, named context-cue-onset asynchrony, or, shortly, onset asynchrony. When the onset asynchrony was set to 0 ms, the cue appeared immediately coloured, either in blue or in red. When, alternatively, onset asynchrony was 300 ms, the cue appeared in black for the first 300 ms, after which it became either blue or red.

### Procedure

Participants in the lab version were tested individually in sound-proof chambers during sessions that last roughly 50 minutes. The experimenter would previously stick an A5 sheet reporting the task-cue and stimulus-response mappings below the computer monitor. They read the instructions and the experimenter ensured these were clear before starting the practice block.

Participants in the online version were sent an email with the link to the experiment (RWTH students), or they accessed the experiment via Prolific. In both cases, they could when they pleased and they had maximum 1 hour, against an average completion time of 25-30 minutes. The information concerning the key-response mapping and the cue-task mapping were presented at the bottom of the screen in each training and experimental trial, see Figure 1.

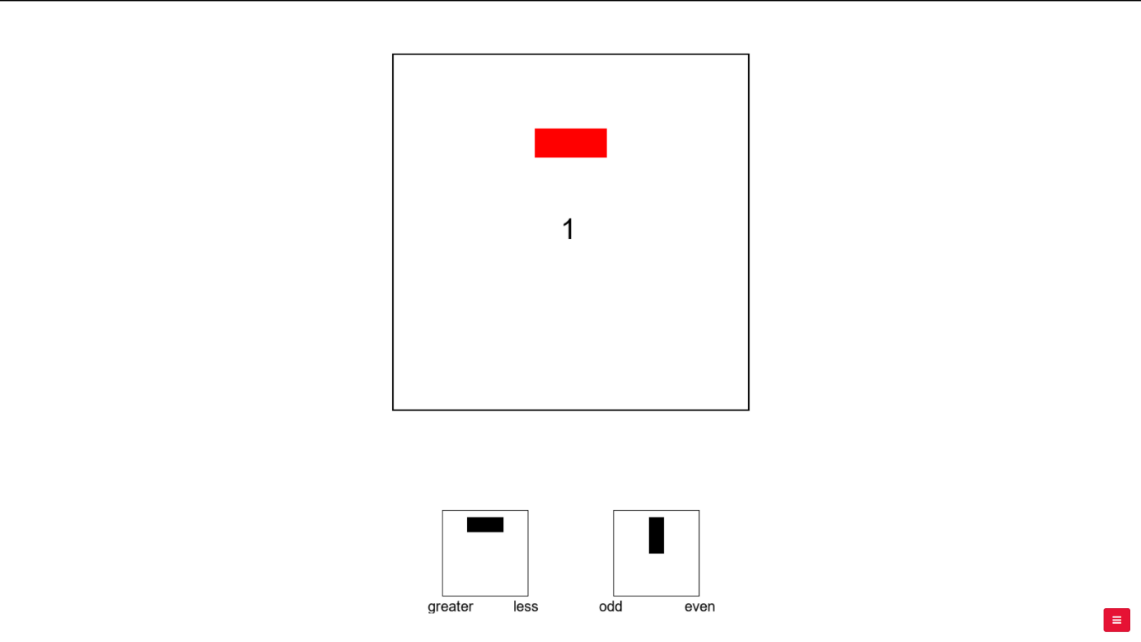


Figure . An example of a trial display at stimulus presentation.

After the instructions, participants performed 16 practice trials, that covered all the possible combinations of stimuli and tasks (8 stimuli x 2 tasks). In this phase, a feedback was given for wrong answers and/or for slow (more than 2 seconds) answers. It consisted in a sentence saying: "The answer is incorrect! Please try to be more careful!" and/or “Please try to be faster!”, which appeared in black on the screen. In the online version, the accuracy feedback was given by means of a green tick or a red cross. During training trials, colour was never added to the cue that was consequently always black.

When laboratory participants declared to the experimenter to be ready, the experiment began, which was composed of 8 blocks of 96 trials. Similarly, also online participants could decide to re-read the instructions and do the training again, or they could decide to start. in order to comply with general recommendations concerning online studies (see, for example, Sauter, Draschkow, & Mack, 2020), we halved the duration of the online experiment, administering 4 blocks instead of 8.

Each trial started with a black fixation cross (1.4 seconds) followed by the simultaneous appearance of the cue and the frame. Both cue and frame remained on the screen for 600 ms in total, after which the stimulus appeared. The three objects remained on the screen until a response was given or a deadline of 2.5 seconds was reached. The following trial started immediately after, with the onset of the fixation cross. No feedback was provided for correct, incorrect, nor slow responses.

Onset asynchrony manipulation was implemented block-wise, so that it was the same each second block and four blocks had positive onset asynchrony (300 ms) and the other four 0 ms onset asynchrony (two and two in the online version). It was counterbalanced across participants whether the blocks with positive onset asynchrony were the even or the odd blocks. Within each tasks sequence in a block, half of the trials were magnitude trials, and half parity trials; the cue was blue in half of the trials and red in the other half, and each digit was presented exactly 12 times. Furthermore, probability of a certain colour given a certain task and a certain stimulus was balanced, so that it was not possible to predict upcoming features of the trial given the other features. Besides, task switch trials were almost the same as task repetitions trials (in each block there was always exactly one task switch more or one task repetition more) and it was never allowed for the same stimulus (the digit) to repeat in subsequent trials.

*Design*

We were interested in testing how the relation of task-irrelevant contextual features in subsequent trials interacted with task and response relation. In other words, we tested whether a n – 1 switch versus a repetition of a task-irrelevant features interacted with a n – 1 task or response switch or repetition. Furthermore, we wanted to assess whether such interaction was modulated by temporal asynchrony between cue onset and context onset. To this aim, we ran a 2x2x2x2 mixed-design analysis of variance (ANOVA), with reaction times[[2]](#footnote-2) as dependent variable and including task relation (task repetition vs. task switch), response relation (response repetition vs. response switch), context relation (context repetition vs. context switch) and onset asynchrony (0 ms vs. 300 ms) as within-subject independent variables, and experiment version (lab vs. online), as a between-subject variable. Furthermore, we ran an identical ANOVA having the average error rate as dependent variable.

## Results

### Analyses of Reaction Times

For this ANOVA, error and post-error trials were removed, together with the first trial of each block and the responses faster than 200 ms: this implied to remove 13.2% of the initial dataset. We found significant main effect of task relation, *F*(1,122) = 161.3, *p* < .001, indicating task switch costs (751 ms vs 852 ms). We found a marginally significant effect of response relation *F*(1,122) = 3.87, *p* = .051, indicating response switch *benefits* (808 ms vs 795 ms), and this entered in a significant interaction with task relation, *F*(1,122) = 28.4, *p* < .001.

Main effect of context relation was significant, *F*(1,122) = 8.02, *p* = .005, indicating context (). Context relation entered in a significant three-way interaction with task relation and response relation, *F*(1,122) = 4.10, *p* = .045. We divided response repetition from response switch trials, and we ran an ANOVA on each subset of the data: results showed that context relation significantly interacted with task relation only in response repetition trials, *F*(1,123) = 7.81, *p* = .006, and not in response switch trials, *F*(1,123) = 0.01, *p* = .908. The interaction in the former subset was driven by task switch costs being reduced in context switch trials with respect to context repetition trials (138 ms vs 113 ms).

Context relation entered in another significant 3-way interaction with task relation and with onset asynchrony, *F*(1,122) = 4.58, *p* = .034. We divided trials with null cue-context onset asynchrony (0 ms) from trials with positive cue-context onset asynchrony (300 ms), and we ran an ANOVA on each subset of the data: results showed that context relation significantly interacted with task relation only when cue-context onset asynchrony was null, *F*(1,123) = 8.63, *p* = .004, and not when it was positive, *F*(1,123) = 0.33, *p* = .565. The interaction in the former subset was driven by task switch costs being reduced in context switch trials with respect to context repetition trials (116 ms vs 87 ms).

Experiment version entered in a significant interaction with response relation, *F*(1,122) = 4.09, *p* = .045, which was driven by the presence of positive response switch benefits in the online dataset, which were instead null in the lab dataset (18 ms vs 0 ms). Experiment version furthermore entered in a significant 4-way interaction with context relation, response relation and onset asynchrony, *F*(1,122) = 4.18, *p* = .043. When data was subset in lab versus online data, context relation, response relation and onset asynchrony entered in a significant three-way interaction in the lab subset only, *F*(1,31) = 4.52, *p* = .042, and not in the online subset, *F*(1,91) = 1.96, *p* = .164. This was due to context relation significantly interacting with response relation only in trials with positive onset asynchrony, *F*(1,31) = 6.4, *p* = .017: while in context repetition trials we observed response switch benefits (770 ms vs 755 ms), we observed response switch *costs* in context switch trials (765 ms vs 775 ms).

1. Among participants that declared their age (4 online participants did not), mean age was not significantly different in the online (24.4 years old) and the lab sample (23.1 years old), *t*(118)= 1.78, *p* = .077; neither was the distribution of sex (50% females in lab sample and 46.7% females in the online one), *χ2*(1) = 0.01, *p* = .906. [↑](#footnote-ref-1)
2. Reaction times distribution was strongly left-skewed in both Experiment 1A and 1B, however ANOVAs on log-transformed Rts yielded the same results as the ANOVAs on raw Rts, therefore the latter are reported for both experiments. [↑](#footnote-ref-2)