**Experiment 1a**

**Method**

**Participants**

Fifty-three members (39 female, mean age 21.8 years) of the University of Edinburgh community participated, paid £8 an hour. All identified English as their native language from birth and had normal or corrected vision with no colour blindness.

**Apparatus**

Stimuli were presented on a 19" CRT monitor, connected to a computer running PsychoPy2 software. A chin rest and mirror stereoscope were positioned 56cm from the monitor, with a vertical divider splitting the display in half so that subjects saw only the left hand side of the screen with their left eye and only the right hand side with their right. Participant responses were given via the cursor keys on a standard keyboard.

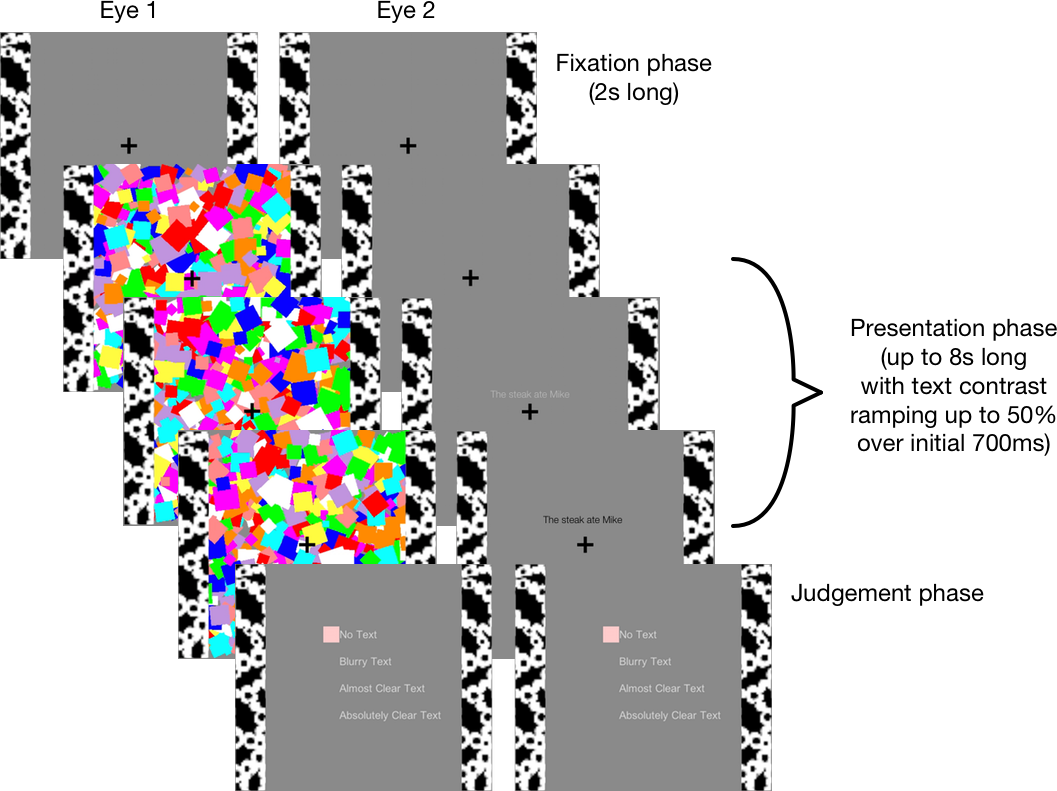
**Procedure**

Each trial consisted of three phases. In the fixation phase, a cross was binocularly presented at the center of each eye's visual field. After 2 seconds, the presentation phase showed one eye a lexical stimulus and the other eye a visual mask (with a fixation cross super-imposed on both). Presentation side was determined randomly for each trial. The lexical stimulus consisted of a sentence either above or below the fixation cross, ramping up from 0% to 50% contrast over 700ms. The position of the text relative to the cross was determined randomly for each trial. The visual mask consisted of a field of squares which randomly changed in size, colour, contrast, rotation and position at a rate of 60Hz. Stimuli were presented in randomised order.

Participants were instructed to focus on the fixation cross with both eyes open, without blinking or looking around, and to quickly press the “up” cursor key if they detected text above the cross, or the “down” cursor key for text below the cross. If the participant did not respond within 8s, the trial timed out. This time-out was not included in Sklar et al.’s original method, but we did not think it was likely to greatly affect our results, as average by-condition reaction times in their experiments were less than one second. As such, reaction times as long as 8s would be considered outliers.

After participants made their judgements, they also reported their subjective experience of the trial. Using a modified version of the perceptual awareness scale (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010), they rated the clarity with which they had experienced the lexical stimulus, choosing from the levels 'No Text', 'Blurry Text', 'Almost Clear Text' and 'Absolutely Clear Text'.

Before beginning the experiment, participants completed five training trials to ensure that the stereoscope was properly calibrated and that they understood the task. Participants were given the opportunity to pause after every 75 trials.



**Figure 1.** Schematic of the experimental procedure.

**Materials**

Lexical stimuli were English translations of the Hebrew expressions used in Sklar et al’s Experiments 1 and 3. These consisted of 34 critical Violation sentences, in which an animate actor performed an implausible action on an object (e.g. *I ironed the coffee*); 68 Control sentences, in which the action or object from the Violation condition was used in a sensible way (e.g. *I made the coffee, I ironed the clothes*); and 34 semantically felicitous Filler sentences (e.g. *I washed the cup*). A full list of stimuli is presented in the appendix.

**Analyses**

Sklar and colleagues used a fairly complex set of criteria for excluding participants and trials from analysis. We followed these criteria in the following order. First, subjects were excluded if their accuracy was below 90%. Second, subjects were excluded if their mean RT was greater than 3 standard deviations away from the grand mean of all subjects. Third, trials were excluded if they were answered incorrectly or timed out. Fourth, trials were excluded for each participant if their RT was greater than 3 standard deviations away from the participant’s grand mean. Fifth, trials were excluded for each condition if their RT was greater than 3 standard deviations from the condition’s grand mean. Finally, trials were excluded if their RT was less than 200ms.