Validating online mouse-tracking software and replicating the disfluency-deception bias

Background

Mouse-tracking

Tracking participants' mouse-movements is a relatively recent method in behavioural psychology. By recording positions and trajectories of the cursor relative to specific responses on a screen, it is possible to study the influence of various experimental manipulations on the decisions participants make during these experiments.
Mouse-tracking has recently gained attention in the context of language research - recording mouse movements alongside the presentation of speech allows researchers to track the time-course of lexical activations during real-time comprehension. For example, Spivey et al.,[1] tasked listeners with responding to spoken instructions such as "Click the candle", while viewing displays which depicted the target (candle) and a distractor. Listeners' mouse trajectories showed a marked attraction towards distractors which shared the same phonological onset with the target word (e.g. candy) compared to distractors which did not (e.g. pickle)

Disfluency and Deception

Previously research has indicated that, when tasked with judging the veracity of an utterance, listeners associate the presence of speech disfluency with dishonesty.
What is more, research from Loy and colleagues[2,3] found evidence of this bias occurring in the early stages of reference comprehension.
Loy et al. (2017) told participants that they were going to hear a speaker describing the location of some hidden treasure. They told participants that the speaker lied about half of the time. Participants viewed a visual display comprising two objects, and heard utterances which specified that the treasure was behind one of the objects in the display. Participants were tasked with clicking on the object *they believed* the treasure to be behind. Crucially, Loy et al. manipulated whether the utterance presented was fluent or disfluent (e.g. "the treasure is behind the/thee - uh - <referent>"). Results showed that participants tended to fixate on and move towards the object not named in the utterance more when the description contained a disfluency than following a fluent utterance.

---

[1] "Continuous attraction toward phonological competitors | PNAS." http://www.pnas.org/content/102/29/10393. Accessed 28 Sep. 2018.
[2] "Effects of Disfluency in Online Interpretation of Deception. - Semantic ...." https://www.semanticscholar.org/paper/Effects-of-Disfluency-in-Online-Interpretation-of-Loy-Rohde/92912b8866c7b98a73df20461e6612fad7e2fc0f. Accessed 26 Sep. 2018.
[3] "Contextual Effects on Online Pragmatic Inferences of Deception ...." 8 Sep. 2017, https://www.tandfonline.com/doi/abs/10.1080/0163853X.2017.1330041. Accessed 26 Sep. 2018.

The Current Project

The study presented here is aimed at validating mouse-tracking software which runs on a server at the University of Edinburgh, allowing the collection of data via online services such as Amazon-Turk[4]. An example of the experiment can be found at: https://shire.ppls.ed.ac.uk/experiment/wKrEBWMCtF

Experiment

We aim to conduct a replication of Experiment 2 from Loy et al.[2] (in which utterance medial disfluency was found to influence judgements of deception) as closely as the mouse-tracking software allows.

Participant recruitment

To allow us to make an explicit comparison between noisy web data and that collected over the web but via a setup that controls for hardware, software, and connection to the server, as well as practical considerations such as visual and audible distractions during the experiment, the experiment will be conducted for two groups.
Firstly, data will be collected from 22 participants recruited from the University of Edinburgh community, and will be conducted in person taking place in an experiment lab for which the above considerations are controlled. This number matches that used in the original study.
The experiment will also be conducted over Amazon-Turk, with the constraint that "mTurk Workers" are located within the United Kingdom, with a planned sample size of 44 participants.

Materials

Visual stimuli consisted of the same 120 line drawings from Snodgrass and Vanderwart[5] that were used in Loy et al.'s study, with the same subset of 20 used in critical trials.
We used the same audio recordings from Loy et al. (Experiment 2) in which the speaker named an object as hiding the location of some treasure. As in Loy et al., the utterances were either fluent or contained a disfluency prior the critical noun:
"the treasure is behind the/thee - uh - <referent>"

Procedure

Figure 1 shows the trial procedure.
A central click box is presented upon which the participants must click (in doing so centering the cursor).
After clicking on this box, the two objects (referent and distractor) appear.
After 1000ms, the playback of the utterance begins.
The participant must then click on one of the objects on the screen, or else the trial will time out 5000ms after the recording ends.

---

[4] "Amazon Mechanical Turk." https://www.mturk.com/. Accessed 28 Sep. 2018.
[5] "Snodgrass (1980) A standardized set of 260 pictures ... - CiteSeerX." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.1979&rep=rep1&type=pdf. Accessed 1 Oct. 2018.

Upon clicking an object (or timeout), the central clickbox is presented, signifying the beginning of the next trial.
Prior to beginning the experiment, instructions (Figure 2) followed by a scoreboard (Figure 3) are presented on screen.

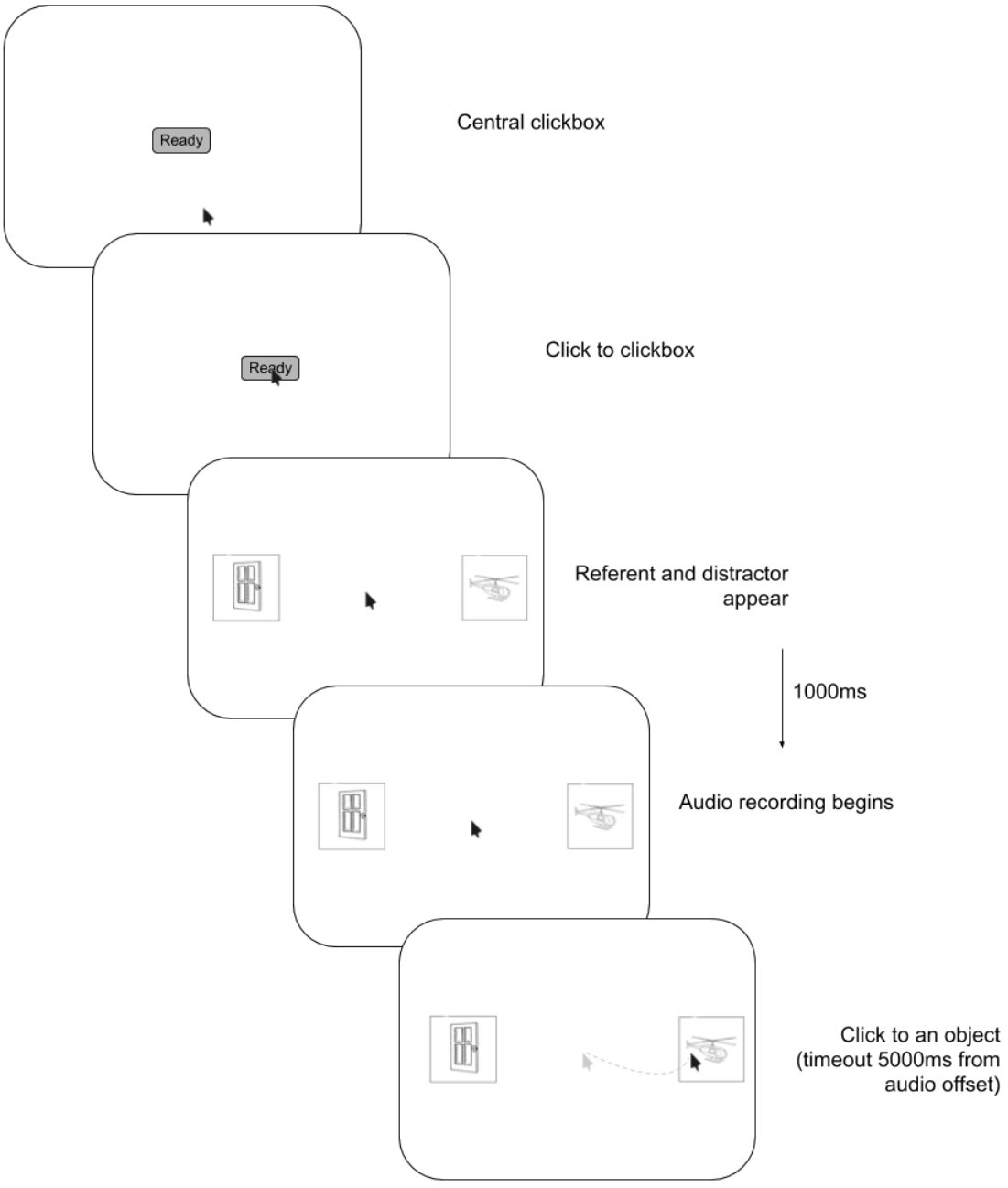

Figure 1. Trial procedure.

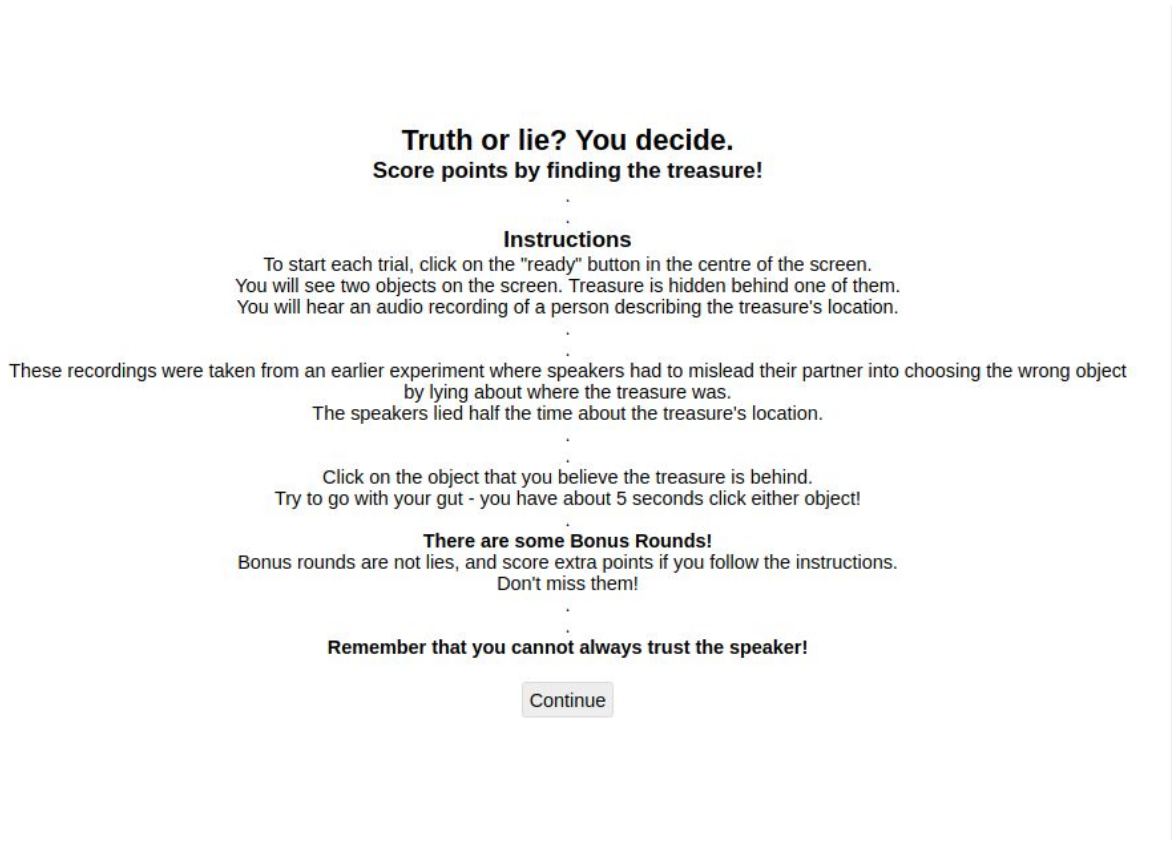

Figure 2. Instruction screen.

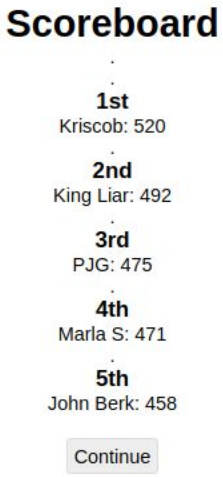

Figure 3. Scoreboard.

Key differences

A number of practical changes were necessary, and are detailed below.

Display, Construction of Materials & Lists, Sample Rates

- Experiment is presented on a 1024x700 display. If the resolution of the monitor is greater than this then the display is centered and the surrounding area is black. The experiment does not run on devices with a lower resolution (e.g. mobile devices). As in the original study, images are centred vertically and positioned such that the center of the images are 15% inwards from either edge of the display.
- In the original study, utterances consisted of two audio clips - the initial sentence fragment and the referent - which were presented consecutively. This made it possible to log messages to the eye-tracker at the appropriate point in time (onset of referent). Playback of the second part of the utterance (the critical noun) occurred at a pre-defined duration after the onset of the first audio clip, and these durations varied across utterances (this was necessary to minimise audio disruption as a result of variability in the experiment script).
  Specifying variable delays to stimuli is not possible with our mouse-tracking software, and playing the 2 audio clips back-to-back with a constant delay/no delay, results in a variable and noticeable delay.
  We therefore concatenated each pair of recordings into one continuous .wav, and the experiment script presents these. The script logs a message at the onset of audio playback, and we can subsequently recenter the data to the onset of the referent by joining the durations of the respective initial sentence fragments (see .csv here)
- Randomisation of trial order and pairings of referents with distractors is predefined for each participants (rather than randomly assigned on each run of the experiment).
- The relative on-screen positions (Left/Right) of the referents/distractors are fully counterbalanced across items. In the original study these were randomly assigned with the constraint that each participant saw referents equally often on either side.
- Mouse position is sampled only when the mouse is clicked or when the cursor is in motion. Preliminary tests suggest that this occurs at a rate of 62.5Hz, or every 16ms (Chrome on Linux with wired connection).

Procedural

- There are no practice trials.
- In the present study, the cursor is always visible, and participants center it themselves in progressing between trials. In the original study, progression between trials required a central fixation point, and the cursor was hidden until utterance onset, at which point it was centered and made visible.
- In the original paradigm, participants are shown a scoreboard at the onset of the experiment, and are told that they will collect points for correctly guessing the true location of the treasure. 25% of filler trials presented participants with feedback stating that the trial was a “bonus round”, and they successfully located the treasure (regardless of which object they clicked).
  Using our software, it is not possible to give feedback on trials.

Instead, we include four “bonus rounds” interspersed randomly throughout the experiment, which serve as attention check trials, to ensure that participants are attending to the experiment and to the audio, and not simply clicking through whilst doing something else.
These bonus rounds are procedurally the same as experimental trials in that they are presented as two objects and along with a spoken utterance.
Each bonus round shows two further images from Snodgrass and Vanderwart (1980), one of which is an animal, and one of which is not. Participants hear an utterance (from a different speaker) saying “This is a bonus round! Simply click on the animal to score extra points!”.
Participants are informed in the instructions that bonus rounds are not lies.

Questionnaire

Prior to the experiment, participants are the following question(s).

1. Was any language besides English spoken in your home before the age of 6?
    a. Yes
    b. No

Analysis strategy

A preliminary data processing and analysis script can be found on the OSF page.

Exclusion criteria

Analysis will be conducted on data from monolingual English speakers. Data from participants who respond to the pre-test question indicating that a language other than English was spoken in their home before the age of 6 will be excluded.

As in Loy et al. (2017), any trials in which participants do not click on either object will be removed from all analyses, as will any trials in which participants click on an object before 200ms after the onset of the referent.

Data from participants who in >40% of critical trials either click on objects before referent onset or fail to click either object will be removed from analyses.

Data from any participants who fail to correctly click on the correct object (the animal) in any of the attention check trials will be excluded.

Mouse clicks

We will analyse the objects (referent vs. distractor) which participants clicked on using a logistic mixed effects model, with a fixed effect of fluency (fluent vs. disfluent, deviation coded), and random intercepts and effects of fluency both by-referent and by-participant (see below).

object_clicked ~ Fluency + (1+Fluency | referent) + (1+Fluency | participant)

Mouse movements

As in the original study, we will calculate the number of pixels moved and the direction of movement (towards either object). Any movements beyond the outer edge of either object will be considered to be `overshooting' and excluded from analyses. For each time bin, we will then calculate the cumulative distance travelled towards either object as a proportion of cumulative distance travelled in either direction from the onset of the referent up until that bin. These proportions will then be empirical logit transformed, and the bias towards the referent over the distractor will be calculated as below:

Elogit bias = elogit(prop. mvmnt to referent) - elogit(prop. mvmnt to distractor)

In this measure, a value of zero indicates no bias towards either object, and positive and negative values indicate a bias towards the referent and distractor respectively. Note that this is different from Loy et al. (2017), in which only movements toward the referent (and not in relation to movements towards the distractor) were modelled.
As in the original paradigm, analysis of critical trials will be conducted on the time window beginning at referent onset and extending for 800ms, just beyond the duration of the longest referent (776ms). Mouse movements will be modelled over this time window using a linear mixed effects model, with fixed effects of time from referent onset (seconds), fluency (fluent vs. disfluent, deviation coded) and their interaction. Random

intercepts and effects for time and fluency will be included both by-referent and by-participant (see below).

Elogit_bias ~ Time * Fluency + (1 + Time + Fluency | referent) + (1 + Time + Fluency | participant)

Following Baayen (2008)[6], we will consider effects in these models to be significant where t > 2.

[6] "Analyzing linguistic data. A practical introduction to statistics." http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf. Accessed 16 Oct. 2018.