eScholarship
University of California

Non-plural interpretations of *some*: Mouse-tracking evidence for quick social reasoning in real time

Wei Li, Psychology, PPLS, University of Edinburgh, UK, s1833067@ed.ac.uk
Hannah Rohde, Linguistics and English Language, PPLS, University of Edinburgh, UK, hannah.rohde@ed.ac.uk
Martin Corley, Psychology, PPLS, University of Edinburgh, UK, Martin.Corley@ed.ac.uk

In support of an account in which disfluency can cue social reasoning in real time, Loy et al. (2019) showed that listeners are more likely to make an early commitment to a socially undesirable meaning of *some* as *all*, if it follows disfluent *uh* in a context where larger values are associated with greed ("I ate, [uh], some biscuits"). However, their finding is also compatible with an account in which disfluency simply heightens attention to the core semantic meaning of *some*, namely, *some and possibly all*. The current study differentiates these two accounts, using contexts in which smaller values are the socially undesirable interpretations of *some*. In two experiments, we recorded participants' mouse movements as they heard fluent and disfluent utterances in a job interview context ("I have, [uh], some qualifications") and clicked on one of four images corresponding to specific interpretations of *some*. Here, in keeping with an account in which the effects of disfluency reflect social reasoning, and contrary to one in which such effects depend on heightened attention, disfluency reduces the value participants associate with *some*. We found that participants were more likely to select images corresponding to one, or zero, qualifications following disfluent utterances. However, their mouse movements show that they are quick to commit to one qualification (Experiment 1) and slow to commit to zero (Experiment 2), suggesting that social context and manner of speech can combine to affect the interpretation of *some* as an utterance unfolds. Assigning its meaning to *one* is relatively easy, but imposing a meaning of zero – in effect, deciding that a speaker is lying – is more demanding.

1. Introduction

When interpreting an utterance in daily conversation, listeners not only take into consideration what is said, but also *how* it is said. For example, disfluencies produced in spontaneous speech, such as fillers like *uh* or *um*, have been shown to affect the inferred meaning of an utterance (Bonnefon et al., 2015; Loy et al., 2017). These inferences are strongly influenced by the (discourse or social) context of the utterances, particularly if a speaker might be thought to be avoiding a hurtful truth, or being deliberately deceptive. However, it is not yet clear how disfluency and context interact to affect a listener's unfolding interpretation of an utterance. In particular, it is not clear whether disfluency has an attentional effect, or whether context interacts with disfluency in processing to affect interpretation more directly. Here, we introduce two web-based mouse-tracking experiments, designed to evaluate the influence of disfluency in context on listeners' online interpretations of ambiguous statements. Our first experiment shows that context does, in fact, interact with disfluency, and our second investigates the nature of that interaction: Whether listeners take disfluency as a sign of speaker deception, or as a sign of a more subtle speaker tactic of communicating a true but strategically selected meaning of an utterance.

The experiments capitalise on the broad ambiguity represented by the scalar quantifier *some*. Consider the following exchange (1) between two speakers:

(1)　　A:　"How many Oreos have you eaten?"
　　　　B:　"I have eaten some Oreos."

Potential interpretations of B's answer include 'I have eaten some and possibly all of the Oreos', the *semantic* meaning, and 'I have eaten some but not all of the Oreos', the *pragmatically-strengthened* meaning. A number of factors have been shown to influence the ways in which individual readers or listeners eventually resolve this ambiguity (Khorsheed & Gotzner, 2023). For example, the lexical alternation between *some* and *some of* can influence listeners' preferences, with the partitive favouring pragmatic strengthening (see Degen, 2015, and references therein). Much of the evidence, however, supports the view that the pragmatically-strengthened meaning is derived via listeners' awareness of the alternative formulations a speaker might have used: Grice's (1975) maxim of Quantity suggests that a speaker would say *all* if that were their intended meaning (Simons, 2017; see Khorsheed et al., 2022, for a review). On this view, the semantic meaning must be calculated first: The listener first accesses what is taken to be the literal meaning describing a subset value or maximal value ('some and possibly all') and subsequently takes into account the availability of the formulation *all* as a speaker's option for specifying the maximal value and, from this, derives the non-maximal 'some but not all' meaning.

To test for a literal-first derivation, Tomlinson, Jr., et al. (2013) asked participants to read sentences like *Some elephants are mammals* and indicate by moving a mouse pointer whether the statements were true or false. Although a majority of participants without prior training

favoured a pragmatic interpretation (i.e., they interpreted the statement as false), their mouse pointers showed initial movements towards the true response, suggesting that *some* was initially interpreted literally. In line with this finding, other studies have shown that participants are slow to preferentially fixate images compatible with the pragmatic meaning of *some* (Huang & Snedeker, 2009), although it has been argued that, in these cases, pragmatic *some* was rendered infelicitous because the experiments also included instructions to fixate images with exact numbers of referents (Grodner et al., 2010).

The findings of Grodner et al. (2010) emphasise the importance of the context in which *some* is encountered. In fact, people's preferences for the pragmatically-strengthened interpretation tend to vary systematically, depending on context: For example, self-paced-reading results show that the 'some but not all' interpretation is favoured when the question under discussion (Roberts, 1996) is "Have all the students passed the exam?" compared to when it is used to answer "How many students have passed the exam?" (Politzer-Ahles & Husband, 2018). With the former question, the focus on whether 'all' is true facilitates the readers' understanding of the pragmatic, 'not all', interpretation of the meaning of *some*.

Besides the linguistic context, information about the social context in which an utterance is produced, or about the speaker, can affect its understanding (e.g., Grodner & Sedivy, 2011; Loy et al., 2017, 2019). On the production side, speakers show a preference to use indirect language when asked to express a meaning amidst competing goals of honesty and kindness; in such contexts, speakers opt for utterances that use scalars to simultaneously avoid being unkind and avoid lying (e.g., "Your poem wasn't terrible" in the face of a mediocre poem; Yoon et al., 2020). Evidence for such behaviour raises the question of whether listeners, in turn, draw inferences about the true situation when they recognize that the speaker has competing goals of needing to be honest while also wanting to stay positive (either out of kindness, as in Yoon et al. (2020), or in order to make a good impression, as in our study). Yoon et al. (2020) and others have used the Rational Speech Act model to capture this interplay between speakers' choices and listeners' inferences when there is ambiguity at play, including with scalar expressions like *some* (Degen, 2023). In another case of pragmatic ambiguity, Beltrama and Schwarz (2024) used a picture-selection task to show that participants were happier to assume that a "laid back, friendly" speaker used values such as $200 to describe prices imprecisely (in this case, to describe an actual value of $212.06) than they were to give "nerdy, geeky" speakers the same latitude. Even where the information about speakers is implicit, listeners tend to associate particular interpretations with particular speakers, for example, deriving the pragmatic meaning of *some* increasingly quickly until a new speaker is introduced to the discourse (Ronderos & Noveck, 2023).

In the case of *some,* social context interacts with the manner of delivery. In a study using written vignettes describing spoken interactions, Bonnefon et al. (2015) suggested that comprehenders could take silent pauses as a cue which shifts expectations towards a less socially desirable meaning. The study manipulated whether or not a silent pause was included

before a speaker uttered a face-threatening expression (see an example in (2) below) and asked participants to judge the speaker's meaning by rating the likelihood that the reply endorsed the socially undesirable meaning (i.e., all people hated your idea).

(2)   Yesterday, you pitched an idea to a group of five persons. Today, you ask Bob (who was in the group) what people thought of your idea. Bob < *stays silent for a few seconds* >. Then he replies: "Some people hated your idea."

The socially undesirable meaning was rated as more likely when the speaker was described as remaining silent before speaking. In this case, participants may infer that the speaker wished to convey to the listener the more charitable 'some but not all' meaning, but their silence suggested that this strategic use of *some* was effortful to produce, requiring additional processing time to access a formulation whose ambiguity affords a face-saving interpretation that could serve to conceal a less desirable 'some and, in fact, all' meaning. Bonnefon et al.'s finding indicates that within a relevant context, the eventual interpretation of *some* can be influenced by the manner in which the speech is delivered.

These effects of disfluency in context hold for online processing of spoken language as well. Loy et al. (2019) used an eye-tracking paradigm in which participants listened to recordings of speakers "confessing" to having eaten from a plate of snacks. In critical trials, a pair of images depicted an empty plate (all snacks eaten) and a plate with a couple of remaining snacks (some, but not all, eaten). Participants were asked to click on the plate which corresponded to the most likely scenario. In trials where the speaker was disfluent ("I ate, uh, some Oreos"), participants were more likely to click on the empty plate, corresponding to the semantically-licensed interpretation that all the snacks had been eaten. Moreover, they were more likely to fixate this plate early, showing that the bias emerged as the speaker's utterance unfolded. Although Bonnefon and colleagues and Loy and colleagues accounted for their findings in terms of social desirability, the tendency to interpret *some* as more likely to mean 'all' could have arisen from a simple attentional bias. Disfluencies have been found to heighten listeners' attention, affecting how listeners process spoken language. Listeners pay special attention to the words preceded by disfluencies and are more likely to subsequently recognise having heard them (Collard et al., 2008; Corley et al., 2007). Beyond the studies on disfluencies, longstanding work highlights the role of attention in error-free lexical activation of target words (Hsu, Kuchinksy, & Novick, 2021; Nozari & Thompson-Schill 2013). If we assume the primary lexical meaning of *some* is its literal semantic meaning before any pragmatic inference is derived, disfluency could serve to focus attention on that semantic 'some and possibly all' core meaning. If this *heightened attention* account holds, the effect of disfluency on listeners' interpretations of the word *some* should tend to favour the semantic interpretation, regardless of the given context.

However, if Bonnefon et al. (2015) and Loy et al. (2019) are correct in that the manner of speech influences listeners' pragmatic interpretation of utterances, we might expect to see people understand *some* differently, depending on the *social reasoning* associated with different contexts. Note that Bonnefon et al.'s (2015) and Loy et al.'s (2019) studies used social contexts where the semantic meaning ('I have eaten some and possibly all Oreos') is also the socially undesirable meaning that a speaker might prefer to conceal, and these studies, therefore, cannot distinguish the attentional and social accounts. The present study introduces a different social context where, critically, the socially undesirable meaning is now the pragmatically-strengthened interpretation of *some*, and this pragmatic 'some but not all' meaning is, thus, expected to be the one that a speaker would prefer to conceal, whereas the 'some and possibly all' meaning remains the core semantic meaning that is potentially focused via the attention-heightening effects of disfluency. The goal is to disambiguate the two posited accounts of listeners' interpretation of *some* (a simple bias arising from heightened attention versus context-sensitive social reasoning).

1.1 The present study

We measured the online interpretation of *some* using a mouse-tracking technique in a web-based task. Specifically, we used as a set-up a series of recorded interviews where participants heard conversations like (3), which manipulated whether the interviewees were disfluent or not when answering ("I've got some 'A's" vs. "I've got, uh, some 'A's").

(3)	Interviewer:	"How many 'A's have you got for your psychology modules?"
	Interviewee:	"I've got some 'A's." (fluent) *or*
		"I've got, uh, some 'A's." (disfluent)

In each of two web-based clicking tasks, participants listened to the recordings while viewing sets of four images depicting different numbers of the relevant qualifications, represented by numbers of 'A' grades, (or numbers of ticks for other qualifications), each time out of a maximum of five, with 'F's (or crosses) representing missing qualifications. Three of the images always represented five, four, or two qualifications. In Experiment 1, the fourth image depicted one qualification, and in Experiment 2, this image depicted zero qualifications. In each trial, participants were asked to click on the image that represented the number of qualifications they thought the speaker held. Even though 'one' is not a usual interpretation of *some*, it maps on to the lower-bound meaning of *some* (i.e., 'at least one'; see Barwise & Cooper, 1981). This value was, therefore, included to test whether it was possible for participants to assign the meaning of *some* to its absolute lower bound ('some but not all, in fact, only one'), given a context that supports the relevant social reasoning. For example, during an interview, an interviewee might use *some* as a "vague exaggeration" of the one 'A' qualification they have actually attained.

We recorded participants' mouse movements throughout the experiment session, measuring for each trial both their clicks (i.e., which image was clicked) and the trajectories of the mouse movements made before each click. If the heightened attention account predicts the real-time processing of *some*, we expect to see results similar to those of Loy et al. (2019), with a preference towards images compatible with the semantic 'some and possibly all' interpretation following disfluency, despite the use of an interview context with a different social bias. However, if listeners' interpretation of *some* is driven by social reasoning, participants are expected to interpret disfluency as indexing the less desirable interpretation of *some*, according to the specific social context. In the interview context used here, a disfluency might lead a listener to assume that the speaker is being deceptive, leading to the assumption that *some* entails a smaller number of qualifications. Accordingly, under this account, we would predict that following disfluent compared to fluent utterances, participants should be less likely to move towards, and click on, images depicting high numbers of qualifications, and more likely to click on those depicting lower values, such as two. Whether the plural expression *some qualifications* can be assigned a value of one is an empirical question.
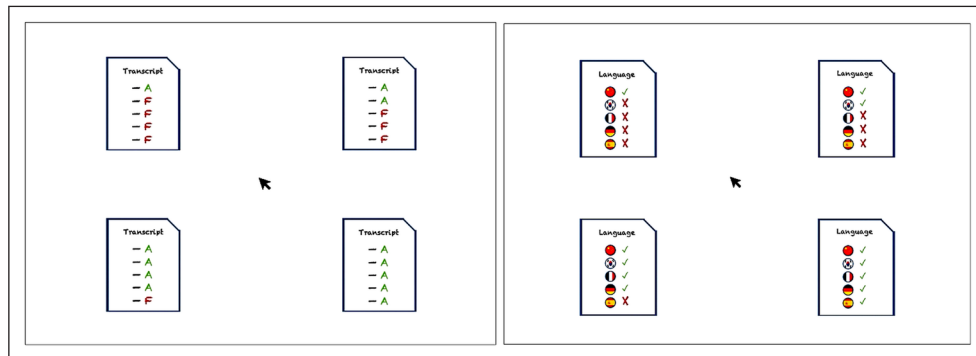
2. Experiment 1

2.1 Methods

2.1.1 Participants

One hundred and fifty (Female: 90, Male: 57, Non-binary: 3) self-reported native English speakers were recruited via Prolific Academic, and each participant was compensated £2 for a task that was estimated to take 15 minutes (£8/h). An additional 29 participants were tested, but their data were not included in our analyses, because these participants (i) failed to finish the whole experiment session, (ii) took more than 40 minutes to finish the whole session (this time limit is based on pre-tests with non-native English speaker participants), or (iii) failed one or both attention checks. All participants were between 18 and 35 years old, with normal or corrected-to-normal vision, and with no self-reported hearing difficulties. Every participant provided informed consent by clicking a button on an information page at the beginning of the experiment, in accord with the University of Edinburgh Research Ethics Committee's approval (reference number: 84/2021-6).

2.1.2 Materials and experiment design

The experiment uses a within-subjects design, manipulating the presence of disfluency (present vs. absent) in critical trials. Each participant listened to 6 interviews, each comprising 6 exchanges between an interviewer and one of 6 interviewees, for a total of 36 exchanges in the experiment. Each interviewee was questioned about the same 6 topics, using the same set of

images (examples for questions on grades and languages are shown in **Figure 1**). Within each interview, 2 of the exchanges were critical trials; one each with a fluent or disfluent response from the interviewee, so that each participant encountered 12 critical trials (6 disfluent). Over the six interviews, each of the 6 types of qualification served once as a critical disfluent item and once as a critical fluent item.

Figure 1: Displays for two target trials in Experiment 1 (Left panel: interviewer asking "How many 'A's have you got for your psychology modules?" and interviewee answering "I've got [uh] some 'A's"; Right panel: interviewer asking "How familiar are you with the listed languages?" and interviewee answering "I can speak [uh] some of them").

2.1.3 Visual stimuli

Six sets of images indicating counts of qualifications corresponding to six interview questions were used as potential referents in the experiment. The six questions concerned the following qualifications, shown in (4):

(4)   Q1: Number of good grades in the psychology modules in the university
            Q2: Number of training courses taken before application
            Q3: Number of criteria the participant meets for the standard of the application
            Q4: Number of languages the participant is familiar with
            Q5: Number of the lab's previous projects the person is familiar with
            Q6: Number of completed recommended reference books

In each target trial, participants saw four images with one, two, four, or all (five) qualifications – number of 'A's versus 'F's (green versus red) or number of ticks marked in green with the remainder marked with a cross in red (as the right panel in **Figure 1**). In one case, green 'A'/ red 'F' was used (as in **Figure 1**). For convenience, these images will be further referred to as *one-tick, two-tick, four-tick* and *all-tick*. In response to each interview question, more ticks or 'A's in the images represent a higher number of qualifications. Images always appeared in fixed locations (one-tick, top left; two-tick, top right; four-tick, bottom left; all-tick, bottom right), to allow participants to learn the response positions and to simplify the mouse trajectory analyses.

2.1.4 Audio Stimuli

For every participant, the whole experiment session included six complete interviews. Every complete interview comprised six conversations between the interviewer and an interviewee, where two of the six conversations were critical trials and the remaining four were filler trials, used to reduce the chance that participants would notice the experimental design. In total, every participant heard 36 exchanges, including 12 critical exchanges and 24 filler exchanges.

Recording of audio stimuli was approved by the University of Edinburgh Ethics Committee with the reference number 119/2021-2. Six native British English speakers (three males and three females) recorded the interviewee scripts, and one native American English speaker recorded the interviewer script.

In critical trials, the recorded interviewees' answers included *some* in either the disfluent or fluent condition. Interviewees' answers in filler trials contained no use of the disfluency *uh* or the word *some*, but included other forms of disfluency (e.g., repetition, prolongation, *like*, *hmph*, etc.) and other expressions of amount (e.g., *almost*, *most of*, *more/less than*, *at least*, *around*, *a few*, *approximately*, *not much*, etc.). For example, the interviewee's answer in a filler trial about the number of training courses taken is "Actually, due to the time limit, I, I, I only took a few of them", using repetition as a disfluency indicator and *a few* to indicate the amount.

The interviewer and interviewee recordings were made online, via a browser-based recording tool implemented in JavaScript.[1] Step-by-step instructions were provided to speakers on how to use the recording website, with details to pay attention to before and during recording. Speakers were asked to record each sentence at least three times, which made it possible for the experimenters to choose the most natural utterances.

The collected recordings were further edited using Audacity (version 2.4.2). In two cases, fluent and disfluent recordings sounded quite different (with obvious intonation differences), and a disfluent *uh* was inserted into a fluent recording; in two cases, the *uh* was excised from disfluent recordings to produce the fluent recordings participants had forgotten to record.

The remaining recordings were all edited in the following way: for fluent utterances, the most natural version was chosen and used; for disfluent utterances, the most natural disfluency *uh* was extracted from one disfluent utterance and was added to the other disfluent utterance (replacing the original *uh*), so that the two critical trials in the same interview had the same disfluency *uh*. For fillers, no editing was applied. In this way, we aimed to standardize our audio recordings without editing out the natural features of speech that tend to occur around disfluencies like *uh*, such as segment lengthening (Shriberg, 2001).

[1] Recordings were made during the Covid-19 pandemic under UK social distancing regulations.

2.1.5 Procedure

Participants completed the experiment, implemented in JavaScript using jsPsych (De Leeuw, 2015), online. Once they had read the welcome page, they clicked the start button and answered questions about age, gender, nationality and languages spoken (these were used to ensure that participants qualified for the study). The main mouse-tracking experiment session started with a trial in which the audio instructed participants to click on a specific image. This was designed to familiarise participants with the experiment, as well as to test that they could easily understand the audio, leaving time for them to adjust the volume as required. Following the audio check, a brief instruction page reviewed the cover story for the study and explained the experimental procedure.

The flow of the experiment is shown in **Figure 2**. At the start of each trial, participants clicked a central Ready button, which ensured that the pointer was at the centre of the screen. After clicking the Ready button, four images appeared on the screen, corresponding to four possible interpretations of the interviewee's answer. After 2000 ms, the first part of the audio (interviewer's question) was played. During this part of the audio, participants could move the pointer using the mouse, but clicks (on any of the four images or elsewhere) were not registered.

Once the interviewee's response started playing, participants were able to click on any of the four images as soon as they wished. If no click was made within 10 s of the interviewee audio offset, the trial timed out and the screen automatically showed the 'Ready' button, indicating the start of the next trial.
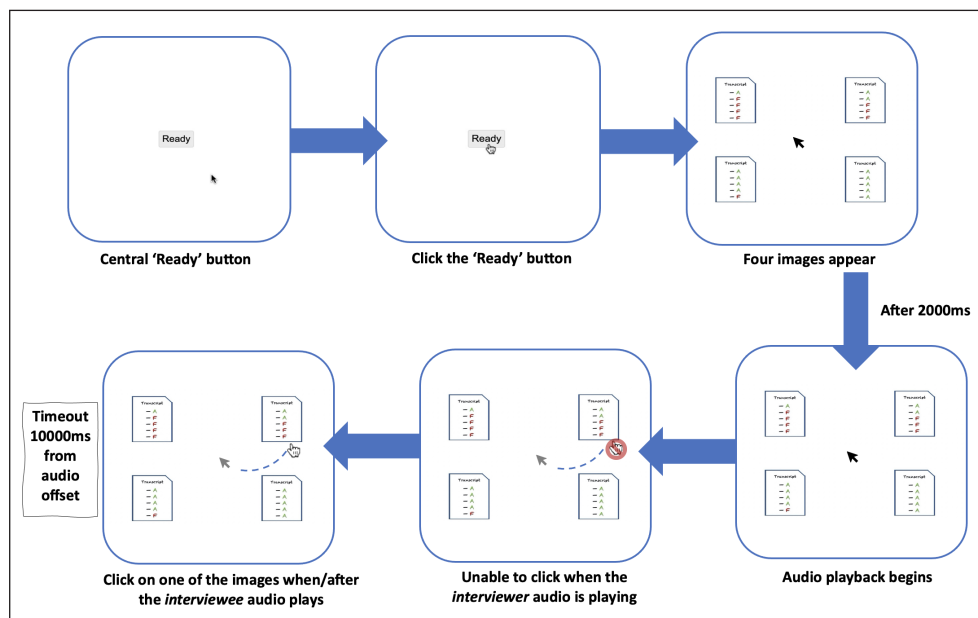
Figure 2: Flow of each trial in the mouse-tracking experiment.

Two attention-check trials were included in the middle (after the third interview conversation) and at the end (after the last interview) of each participant's session. In these trials, the interviewer gave straightforward information indicating which image should be clicked. For example, in the middle attention check trial, the interviewer said, "Among the candidates we have interviewed so far, I'm thinking about ruling out the one who knows the least about our lab", such that the interviewer question uniquely determined the answer (left-bottom image in **Figure 3**).

Figure 3: Attention-check trial example (mid-attention-check).

The mouse-tracking stage finished with a second attention-check trial, after which participants were directed to a Qualtrics questionnaire page, where they answered questions about the experiment and were asked if they had noticed what the experiment was about.

For each experimental item, we recorded the mouse pointer position over the time as well as the identity of the object clicked.

2.2 Results

Data and analysis scripts are available at https://osf.io/95fwh. Prior to analysis, data from filler trials, as well as practice and attention-check trials, were removed. All the statistical analyses were carried out in R version 4.4.1 (R Core Team, 2024).

2.2.1 Final object click

Total numbers of clicks on each image by condition (disfluent/fluent) are shown in **Table 1**. For the disfluent condition, compared to the fluent condition, there were fewer clicks on two- and four-tick images, but more clicks on the one-tick image. Clicks on the one-tick image (coded as one; clicks on any other image, coded as zero) were modelled using a mixed-effects logistic regression, with one fixed effect for the within-participant and within-item predictor, the manner of delivery (fluent vs. disfluent), and random effects with by-participant and by-item random intercepts and slopes. Participants were more likely to click on the one-tick image following a disfluent utterance compared to a fluent utterance ($\beta = 1.185$, SE $= 0.494$, $p < 0.05$). This response pattern is in keeping with the predictions for the social reasoning account, suggesting that the presence of disfluency does, indeed, bias interpretation in favour of the socially undesirable meaning, here, the pragmatic 'some but not all' interpretation of *some*.

Table 1: Total number and distribution of mouse clicks recorded on each image (one-tick, two-tick, four-tick, or all-tick) by manner of delivery (disfluent/fluent).

	One-tick	**Two-tick**	**Four-tick**	**All-tick**
Disfluent	148 (17%)	688 (78%)	45 (5%)	0 (0%)
Fluent	67 (8%)	724 (83%)	78 (9%)	1 ($\approx$ 0%)

2.2.2 Mouse movements

The mouse-tracking data obtained from all 150 participants were processed prior to conducting the analysis. For each trial for each individual, mouse trajectories were time-normalised into 101 time-steps (following Dale et al., 2007, and Spivey et al., 2005). Step 0 represents the onset of the interviewee audio and step 100 corresponds to the time point when the participant made the click. All movements were calculated relative to an origin (0,0) which represented the centre of the screen.

Participants' mouse trajectories towards each image in each condition (disfluent/fluent) are shown in **Figure 4**. Each line represents the average trajectory of the mouse pointer, starting at the onset of the interviewee's response and ending when an image is clicked. Coloured points indicate relative time. The colours (from red points at the centre of the screen to violet in each corner) indicate each 10% of trajectory time.

The all-tick image was only selected in one trial. As can be seen in **Figure 4**, the pointer trajectories for two- and four-tick images show similar patterns in each condition, both in terms of path and in terms of distance over time. However, the patterns of mouse movements
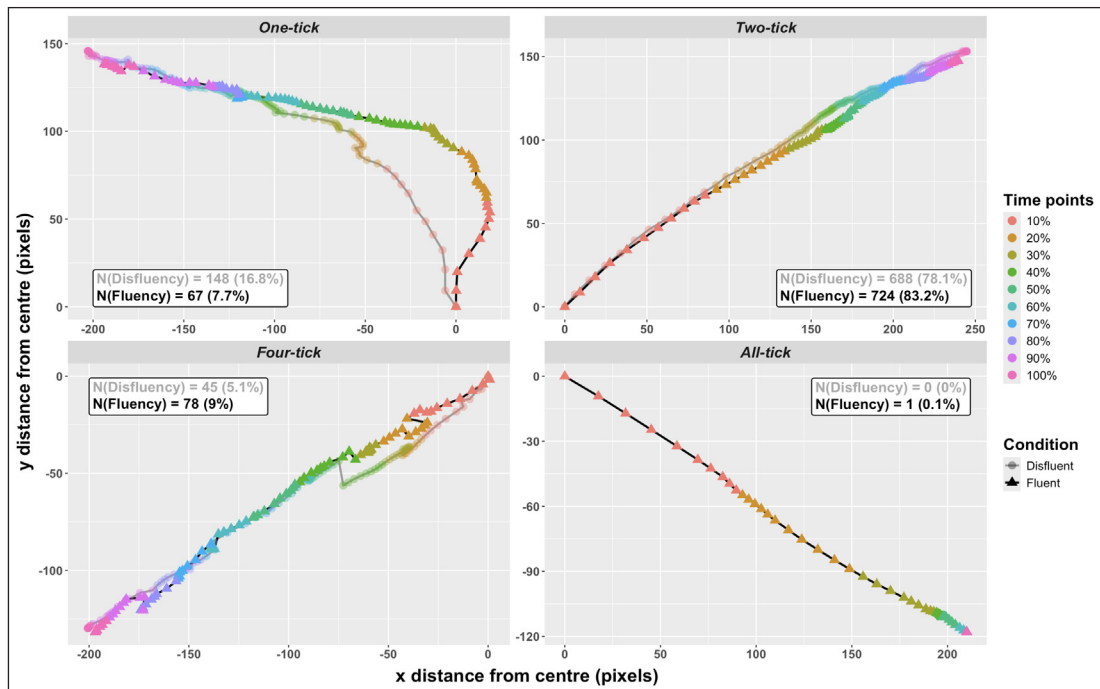
Figure 4: Aggregated mouse trajectories towards four images (one-, two-, four-, all-tick) by condition (disfluent/fluent), and the colours of points (from red points at the centre of the screen to violet in each corner) indicate 10%, 20%, 30%…100% of trial time. Note that the disfluent condition is represented with the grey line with round points in relatively lower saturation, and the fluent condition is represented with the black line with triangle points in normal saturation.

in which participants eventually click on the one-tick image show obvious differences between conditions. Participants appear to move faster and more directly towards the one-tick image in the disfluent condition. In the fluent condition (the less saturated line in **Figure 4**), they make hesitant movements towards other images, taking more of the period between the onset of the interviewee's response and their own mouse click to make a decision to move towards, and eventually click on the one-tick target.

2.2.3 Mouse trajectory analysis

To further explore differences in pointer trajectories between disfluent and fluent conditions, we used the Bootstrapped Differences of Timeseries (BDOTS) package (version 1.2.5; Nolte et al., 2023) in R to analyse our mouse-tracking data. BDOTS is a statistical method which is designed to estimate the time window during which two time-curves differ (see Oleson et al., 2017; Seedorff et al., 2018, for details). Its ability to detect subtle differences in the temporal dynamics of mouse movements makes it suitable for investigating real-time processing.

To perform this analysis, we created a distance variable, which represents the perpendicular distance from each mouse point to the diagonal line formed by joining the centres of the two- and four-tick images, henceforth referred to as the *two-four diagonal line* (see **Figure 5**). We chose this diagonal line as a reference line because the two- and four-tick images both represent meanings that are compatible with the pragmatic (some-but-not-all) meaning of *some* and are the most clicked choices. The perpendicular distance from this line, therefore, depicts the extent to which a participant moves away from common, plural, understandings of *some*. In this sense, analysing this distance over time allows us to map listeners' *non-plural tendencies* when hearing disfluent/fluent utterances.[2]
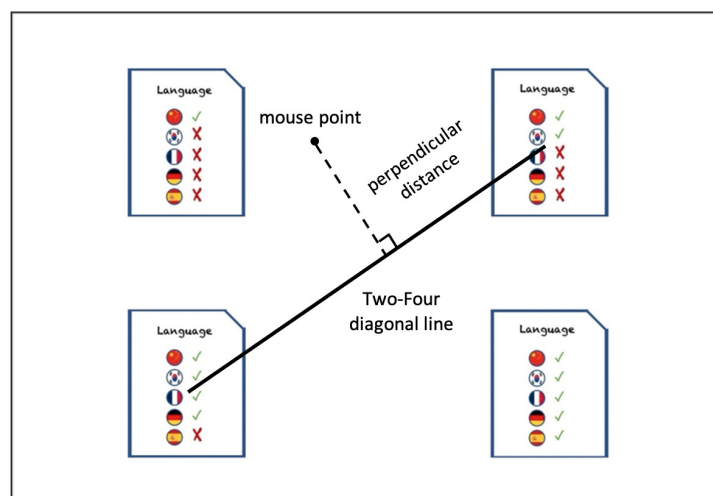
Figure 5: A visual representation of the perpendicular distance from a given mouse point to the diagonal line formed by two- and four-tick images.

Each participant's average mousetrack in each condition was fit using a 4-parameter logistic function. These functions have previously been used to model the increasing likelihood, over time, of fixating a target in visual world eye-tracking experiments (Oleson et al., 2017); our distance metric increases in a similar way. Curve parameters were bootstrapped (1000 iterations) and averaged, allowing us to compare the curves for disfluent and fluent conditions (top panel in **Figure 6**), regardless of where the final clicks were made. In the fitting stage, 8 curves had good fits with autocorrelation between consecutive steps (AR1: $R^2 >= .95$), 20 curves had good fits without AR1, 1 curve had reasonable fits with AR1 ($R^2 >= .80$) and 60 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.994, and the adjusted alpha was calculated to be 0.0102. We found regions of significance starting from time step 22 to the end of the trial time (bottom panel in **Figure 6**).

---

[2] We appreciate that movements towards items at the bottom right (all-tick) represent a plural interpretation of *some* (as 'all'). However, since all-tick was only selected in one instance, *non-plural* seems to be the most useful term to use for expositional purposes.
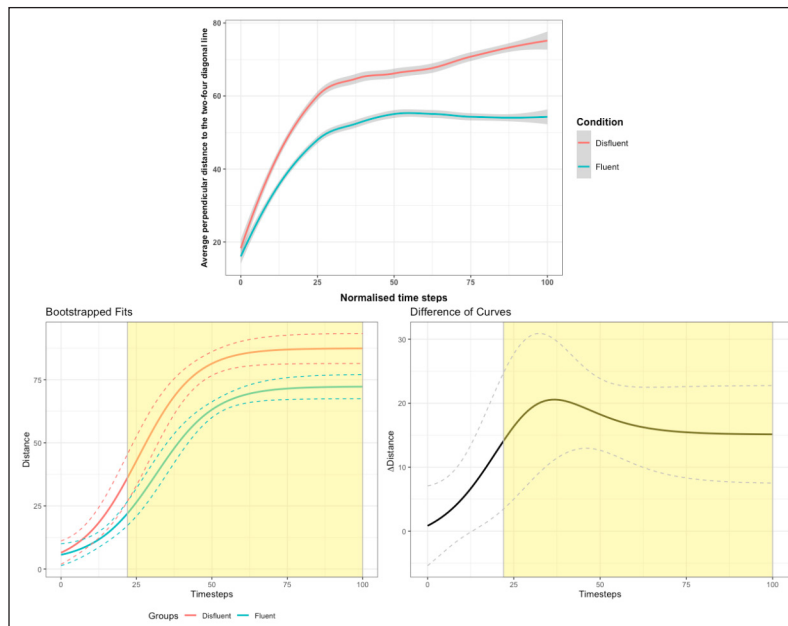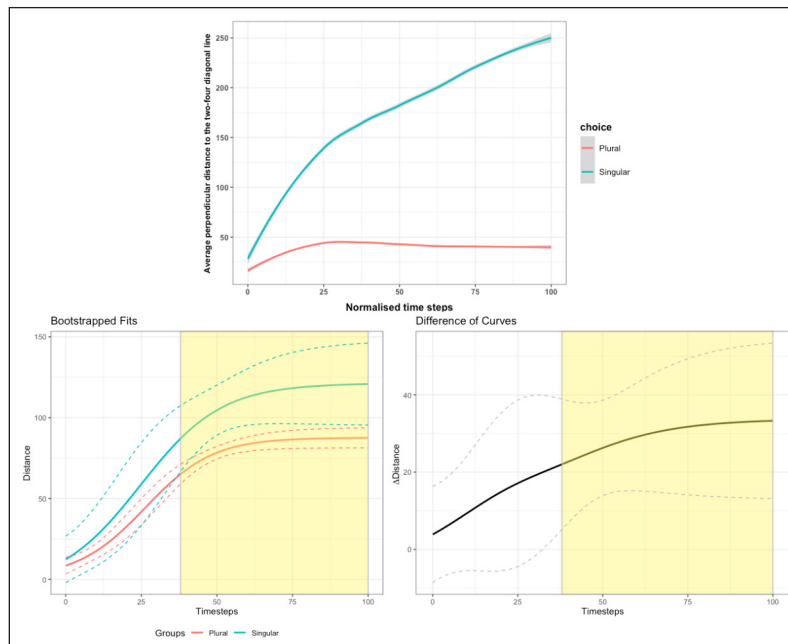
Figure 6: *(top panel)* Two time-curves showing average perpendicular distance (in pixels) from the mouse pointer to the two-four diagonal line across 101 time steps by disfluency conditions; *(bottom panel)* BDOTS results of curve differences between conditions (disfluent/fluent), highlighting the significant time period of differences in Experiment 1.

The BDOTS analysis shows that participants were faster to move away (in perpendicular distance) from the two-four diagonal line in the disfluent, compared to the fluent, condition. It indicates that following disfluent utterances, listeners are more likely to select a non-plural meaning of *some* (higher asymptote) and to make that decision quickly (steeper slope). In particular, differences between conditions emerge at step 28 in the mouse trajectory.

In order to investigate further the decision to click on the one-tick image, we isolated the disfluent condition and compared trials in which one-tick was eventually selected to those in which it was not. By-participant data in the disfluent condition were again fit using a 4-parameter logistic function. We compared the distance metrics between those clicking on the one-tick image (noted as *singular* image) and those clicking on other images (*plural* images) (top panel in **Figure 7**).

In the fitting stage, 14 curves had good fits with AR1 ($R^2 >= .95$), 31 curves had good fits without AR1, 8 curves had reasonable fits with AR1 ($R^2 >= .80$) and 91 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.996, and the adjusted alpha was calculated to be 0.0124. We found regions of significance starting from time step 38 to the end of the trial time (bottom panel in **Figure 7**).

Mapping the results within the disfluent condition onto actual timings of mouse movements shows that the non-plural tendency emerged quite quickly: When participants encountering disfluency decided to click on the one-tick image, their mouse movements diverged around 474

Figure 7: *(top panel)* Two time-curves showing average perpendicular distance from the mouse pointer to the two-four diagonal line across 101 time steps by where the final click is singular (one-tick) or plural (two-, four-, all-tick) images; *(bottom panel)* BDOTS results of curve differences between final clicks on singular option (one-tick image) and plural options (two-, four-, all-tick images), with the significant time period of differences highlighted and average disfluency onset noted (at around step 30) in Experiment 1.

ms after average disfluency onset. This suggests that individuals decided to click on the one-tick image almost as soon as they heard the disfluency.

2.3 Discussion

Experiment 1 builds on prior work which has established effects of disfluency on the interpretation of *some*, but distinguishes between a heightened attention and a social reasoning account of these effects. Specifically, we tested whether disfluency preceding *some* caused listeners to pay increased attention to the literal meaning of 'some and possibly all', or whether listeners tended to interpret *some* in a social context, with disfluency heightening the salience of a socially undesirable interpretation.

Experiment 1 supports the social reasoning account in two ways. First, in a context where it is desirable to have more qualifications, disfluency biases towards socially undesirable pragmatic interpretations of *some* which imply smaller values. In fact, it would appear that the meaning of *some* is at its lower bound, in that participants accepted an interpretation where *some* was associated with the meaning 'one' in 17% of disfluent and 8% of fluent trials. Taken together with the findings from Loy et al. (2019), this suggests that listeners take the social context into

account when reasoning about scalar quantifiers, such as *some*. Second, this reasoning happens very quickly: Where utterances are disfluent, listeners who are going to choose the one-tick image make the decision to select that target within 355 ms, and initiate their mouse movements appropriately. Where the utterances are fluent, they are more hesitant, making early mouse movements which suggest that the two-tick interpretation of *some* is initially in contention.

However, even though the value of one is compatible with *some* (in its 'at least one' meaning), interpreting *some* as meaning 'one' is still an unexpected finding; listeners typically associate *some* with a plural meaning, given the availability of the word *one* to express the more specific lower-bound value (see also Degen & Tanenhaus, 2015). This leaves open an alternative interpretation, which Experiment 2 aims to explore.

3. Experiment 2

Experiment 1 shows that social reasoning can affect the ways in which listeners interpret the meaning of *some*. However, the surprising finding that clicks on one-tick targets became more frequent following disfluent utterances raises the question of how that meaning emerges. On the one hand, comprehenders could be using what we will term a *lower-bound* strategy: Disfluency in context causes listeners to associate the meaning of *some* with the lowest possible value compatible with the meaning of *some* ('at least one'). On the other hand, listeners might reason that an expression like *some qualifications,* despite its plural marking, does not convey a plural value at all in the given context, perhaps because disfluent filled pauses, specifically, *um* and *uh,* have been shown to act as cues to deception (DePaulo et al., 1982; Loy et al., 2017, 2018). If the listener infers that the speaker is being deceptive, they may be able to ignore the meaning of *some* entirely (the *overwriting* account).

To distinguish these two possibilities, Experiment 2 replicated Experiment 1 with one change to the design. In Experiment 2, the one-tick images were replaced with zero-tick images. The value zero is not compatible with any meaning of the word *some*. If listeners are still making interpretations based on the meaning of *some,* as supposed by the lower-bound account, we would not expect to see clicks on the zero-tick image. However, if listeners are abandoning a plural interpretation of the expression *some Xs* when speakers in this experiment are disfluent, they may click on the zero-tick image.

3.1 Methods

3.1.1 Participants

We recruited participants and filtered collected data in the same way as for Experiment 1. Mouse movement data from one hundred and seventy-three (Female: 87; Male: 82; Non-binary: 4) self-reported native English speakers recruited via Prolific Academic were analysed further, with data

from an additional 11 participants ruled out (either because they took more than 40 minutes to finish the whole session, or they failed one or both attention checks). Participants provided informed consent by clicking the "I consent" button on a consent page, following the University of Edinburgh Research Ethics Committee guidelines (reference number: 84/2021-8).

3.1.2 Material and experiment design

As in Experiment 1, we recorded participants' mouse movements in a web-based task. Participants heard recorded interview conversations between an interviewer and one of six interviewees, in which we manipulated Disfluency (present vs. absent) within subjects in a set of target trials (N = 12). For each target trial, participants saw four images with four different counts of qualifications displayed on the screen (see **Figure 8** for example trials). One-qualification images used in Experiment 1 were replaced with zero-qualification images in the present experiment. Experiment 2 was implemented in jsPsych version 7.0 (De Leeuw, 2015).
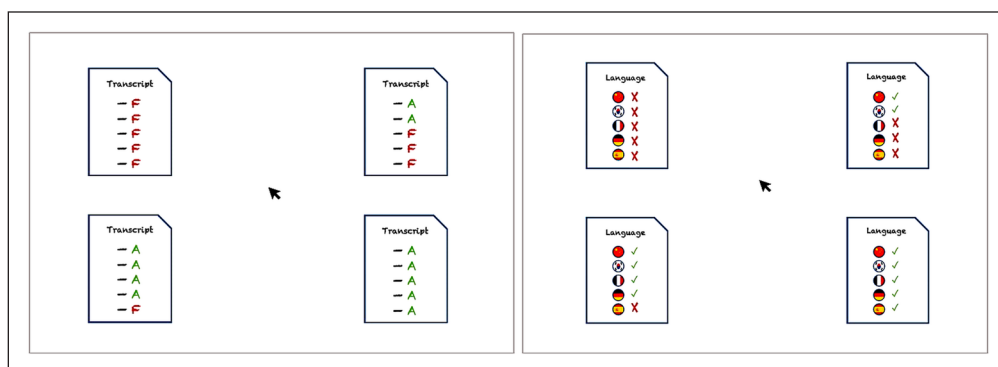
Figure 8: Displays of two target trials in Experiment 2, the same as Experiment 1, except that the top left in both panels consists of a zero-qualification image.

3.1.3 Audio and visual stimuli

The audio stimuli from Experiment 1 were used in Experiment 2. For the visual stimuli, participants saw four images from one set with zero, two, four or all (five) qualifications marked with green ticks and red crosses (green 'A' and red 'F' in one case). As in Experiment 1, the images always appeared in fixed locations (with zero-tick on the top left; two-tick, top right; four-tick, bottom left; and all-tick, bottom right).

3.1.4 Procedure

Other than the use of zero-tick, in place of one-tick, images, the procedure was identical to that for Experiment 1.

3.2 Results

We analysed the data from Experiment 2 in the same way as for Experiment 1, in R version 4.4.1 (R Core Team, 2024). Filler trials, practice trials and attention-check trials were removed prior to analysis.

3.2.1 Final object click

Table 2 shows the total numbers, as well as the percentages, of clicks on each image by condition (disfluent/fluent). Similar to the clicks in Experiment 1, participants' responses suggest a general bias towards a non-plural option, cued by disfluency: Following disfluent utterances, compared to fluent utterances, people clicked more on the zero-tick image (19% vs. 6%) and less on two- and four-tick images (81% vs. 93%). As in Experiment 1, where clicks on the one-tick image increased by 10% following disfluency, participants' choices are influenced by speakers' manner of speech ($\beta = 1.899$, SE $= 0.42$, $p < 0.001$). This response pattern is in keeping with the predictions of the *overwriting* account, suggesting that the presence of disfluency allows participants to consider the zero-tick interpretation of the interviewee's utterance, which is incompatible with standard meanings of *some*.

Table 2: Total number and distribution of mouse clicks recorded on each image (zero-tick, two-tick, four-tick, or all-tick) by manner of delivery (disfluent/fluent).

	Zero-tick	**Two-tick**	**Four-tick**	**All-tick**
Disfluent	191 (19%)	782 (78%)	35 (3%)	0 (0%)
Fluent	62 (6%)	883 (87%)	65 (6%)	1 (1%)

3.2.2 Mouse movements

We processed the collected mouse-tracking data from all 173 participants, following the same two key steps as in Experiment 1. Although participants appeared willing to interpret *some* as 'zero', inspection of their mouse movements suggests that the processes underlying this decision differ from those which underlie an interpretation of *some* as meaning 'one'. Figure 12 shows participants' average mouse trajectories towards each image for fluent and disfluent items. The colours of the points (from red to violet) each indicate 10% of the trial time. We consider the trajectories for fluent versus disfluent conditions in the cases where participants chose either the one-tick interpretation (Experiment 1) or the zero-tick interpretation (Experiment 2). When ultimately clicking on the one-tick image (top-left panel in Experiment 1, see **Figure 5**), participants show hesitant and less direct movements when the utterance is fluent. However, when ultimately clicking on a zero-tick image (top-left panel in Experiment 2, see **Figure 9**), participants show hesitant movement towards other images in both conditions.
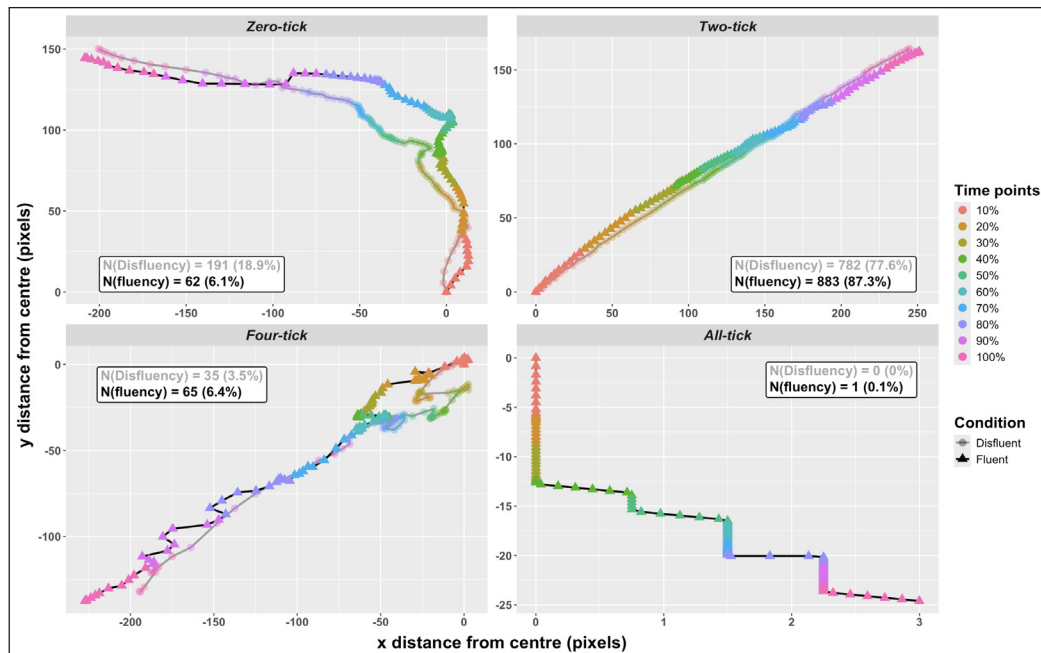
Figure 9: Aggregated mouse trajectories towards four images (zero-, two-, four-, all-tick) by condition (disfluent/fluent), and the colours of points (from red points at the centre of the screen to violet in each corner) indicate 10%, 20%, 30%…100% of trial time. Note that the disfluent condition is represented with the grey line with round points in relatively lower saturation, and the fluent condition is represented with the black line with triangle points in normal saturation.

3.2.3 Mouse trajectories analysis

Again, we calculated the perpendicular distance between mouse pointers to the two-four reference line to indicate the non-plural tendency. Obtained values were fit using a 4-parameter logistic function in a BDOTS analysis. We first compared the difference of time-curves for disfluent and fluent conditions across all items (top panel in **Figure 10**).

In the fitting stage, 10 curves had good fits with AR1 ($R^2 > = .95$), 28 curves had good fits without AR1, 12 curves had reasonable fits with AR1 ($R^2 > = .80$) and 65 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 1 and the adjusted alpha was calculated to be 0.0332, showing that we failed to find any significant regions where these two time-curves start to differ. This result (bottom panel in **Figure 10**) suggests that in this experiment, where zero-tick replaces the one-tick image, manner of speech does not significantly influence the timing of participants' choice towards which image to click on.

As for Experiment 1, we performed a second analysis, focusing on the disfluent condition. We fit the data by participants using a 4-parameter logistic function, to compare the curve differences between participants clicking on the zero and those clicking on plural images (top panel in **Figure 11**).
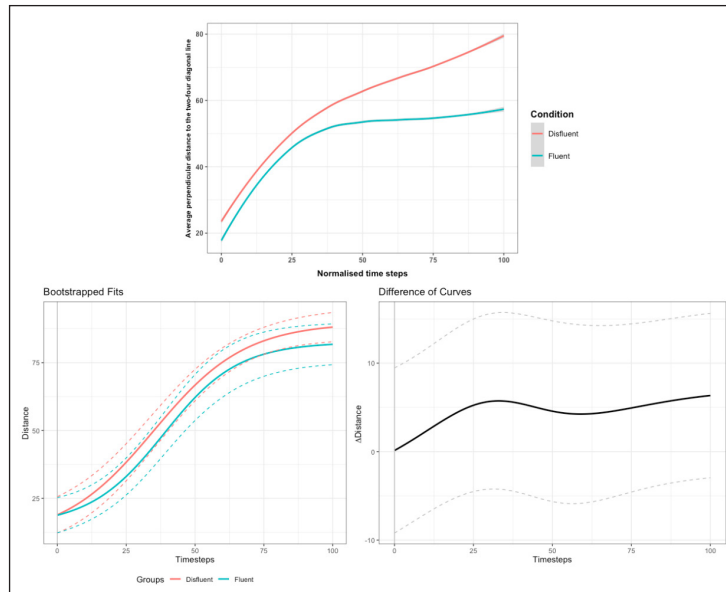
Figure 10: Two time-curves showing average perpendicular distance (in pixels) from the mouse pointer to the two-four diagonal line across 101 time steps by disfluency conditions (top panel); BDOTS results of curve differences between conditions (disfluent/fluent), highlighting the significant time period of differences in Experiment 2 (bottom panel).
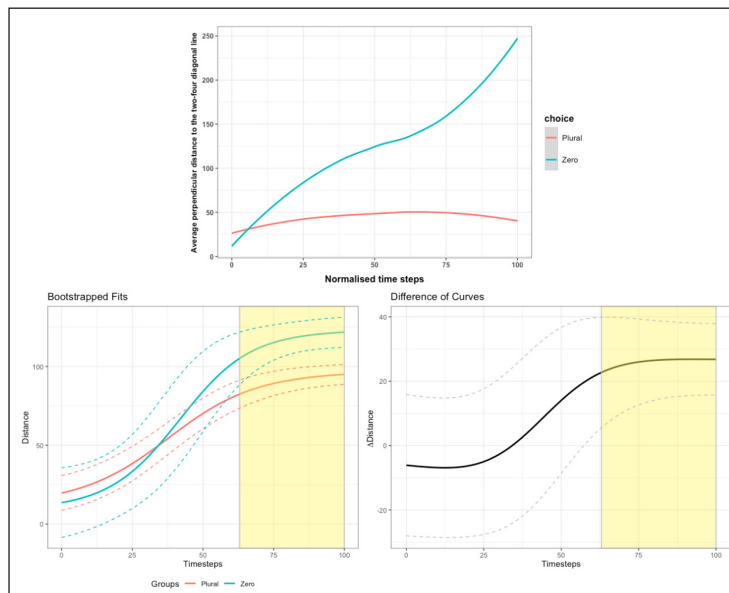
Figure 11: *(top panel)* Two time-curves showing average perpendicular distance from the mouse pointer to the two-four diagonal line across 101 time steps, where the final click is zero (zero-tick) or plural (two-, four-, all-tick) images; *(bottom panel)* BDOTS results of curve differences between final clicks on the zero option (zero-tick image) and plural options (two-, four-, all-tick images), with the significant time period of differences highlighted and average disfluency onset noted (at around step 36).

In the fitting stage, 24 curves had good fits with AR1 ($R^2 >= .95$), 38 curves had good fits without AR1, 19 curves had reasonable fits with AR1 ($R^2 >= .80$) and 90 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.996 and the adjusted alpha was calculated to be 0.0126. We found regions of significance starting from time step 63 to the end of the trial time (bottom panel in **Figure 11**).

When participants encountered disfluent utterances, the mouse movements in the trials where they clicked on the zero-tick image diverged at around step 63 from those in the trials where other images were clicked. If mapping this time onto actual mouse movement timings, this indicates that participants decide to choose the zero-tick image over the other ones around 1357 ms after average disfluency onset. For comparison, the equivalent value for Experiment 1 was 474 ms.

4. General discussion

The current study explores how listeners' interpretations of the ambiguous scalar quantifier *some* are influenced by the speaker's manner of speech. Previous research has suggested that in deriving pragmatic meaning, readers or listeners bring together the social context, what they know about the speaker, and even the types of inference the speaker's previous utterances have supported (Beltrama & Schwarz, 2024; Loy et al., 2017, 2019; Ronderos & Noveck, 2023); here, we show how disfluency interacts with this social context, investigating whether participants reason socially about the fluency with which utterances including the word *some* are produced.

Different from the context used in Loy et al. (2019), where larger values are socially undesirable ("I ate, uh, some Oreos"), we set up a social context where the undesirable meaning corresponds to smaller values of *some* ("I've got, uh, some 'A's"). Our experiments show that in this 'job interview' context, disfluency has the opposite effect from that shown in Loy et al.'s study, reducing the values participants associate with *some*: We found that participants are more likely to select images corresponding to one or zero qualifications following disfluent utterances. These results complement the findings of Loy et al. (2019), establishing that both results likely stem from participants' social reasoning about the presence of disfluency in a context with *some*.

The increase in numbers of clicks on the lowest available number of qualifications is comparable across experiments. Participants' mouse movement patterns show that, following disfluency, they are relatively quick to commit to one (Experiment 1) or zero qualifications (Experiment 2) when these are the interpretations they choose, with differences emerging shortly after the onset of the disfluency. By contrast, the trajectories of participants' mouse pointers appear to differ qualitatively between experiments (**Figures 4** and **9**), emphasising the point that similar final-decision outcomes may be achieved by different cognitive processes (e.g., Li et al., 2021).

A limitation of the experiment is that our comparison of timeseries does not allow for a direct comparison of the times at which trajectories diverge in Experiments 1 and 2. We note, however, that participants' trajectories tended to move further from the plural line (i.e., closer to one qualification) following disfluency in Experiment 1, but moves towards zero qualifications were not similarly affected in Experiment 2. Moreover, on a purely numeric basis, our analyses suggest that a decision to click on zero qualifications may emerge later than a decision to click on one qualification, raising the interesting possibility that accessing the lower-bound meaning of *some* as 'one' is relatively easy, but overwriting it with 'zero' (in effect, deciding that a speaker is lying) may be more demanding. We emphasise, however, that these observations are not based on a statistical comparison of the two experiments, and should, therefore, be treated with appropriate caution.

To the extent that the difference between experiments can be interpreted, a potential account is that both the lower-bound and overwriting strategies are in play when listeners are searching for an interpretation of ambiguous *some*. A context such as that of an interview may set up an expectation that speakers could potentially be misleading or deceiving. Bearing possible deception in mind, listeners who encounter disfluency try to push the boundary of the meaning of *some* as low as possible. When the lowest value available is still within the boundary of the semantic meaning of *some* ('one'), listeners seem to arrive at this associated meaning relatively quickly. However, when the available lowest value is not included in the meaning of *some* ('zero'), listeners hesitate before overwriting the interpretation to zero, if they are going to choose that option. In lay terms, it is easier to assume that someone is exaggerating than to assume that they are lying. Alternatively, a lack of difference between experiments would be compatible with the strong claim that implicatures are not necessarily calculated when people assess the likely meaning of *some* (Kissine & De Brabanter, 2023).

Whichever of these accounts is correct, our experiments suggest that, in the right circumstances, participants can relatively easily ignore the lower-bound semantic meaning of *some*. This raises a question concerning studies which have varied the salience of 'all', the upper-bound semantic meaning of *some* (e.g., Breheny et al., 2013; Breheny et al., 2006; Politzer-Ahles & Husband, 2018). Although 'all' is compatible with a semantic interpretation of *some*, it is also feasible that a pragmatic interpretation is simply overwritten (perhaps with literal *all*) in the relevant circumstances. Explorations of the upper-bound semantic meaning are unlikely to distinguish these accounts (especially without information on the timecourse of interpretation).

Now that the possibility of overwriting has been raised, and demonstrated, it would appear that investigating lower-bound implicatures of *some* may be more profitable, allowing us to move beyond the relative availability of the 'some but not all' implicature to ask a more general question about how listeners use what a speaker says to recover their social goals. Scalar *some* provides a convenient probe, because of its ambiguity, but future studies need not be limited to

this particular interpretational ambiguity. The findings we present here suggest that listeners are sensitive to the possibility that speakers may make creative and strategic use of the linguistic forms available to them in trying to cooperatively – and not so cooperatively – guide the communicative interaction.

Methodologically, our current study shows the possibility of using a web-based mouse-tracking task to explore real-time processing, with the advantage of easy implementation and quick recruitment of participants. With four images on the screen, our materials were more complex than those in previous mouse-tracking studies (with two images located symmetrically). As such, our study confirms the viability of studying more complex visual scenes and of using visualisation and time-course analysis of trajectories that go beyond standard methods of analysis for mouse-tracking data (Spivey et al., 2005).

Our two experiments suggest that manner of speech facilitates listeners' understanding of the scalar quantifier *some* via a process of social reasoning about speakers' goals, and that this reasoning happens fast, in the early stages of listeners' comprehension. Given the role of social reasoning, these findings open up new questions about possible differences in this kind of reasoning across individuals with varying pragmatic abilities or across contexts that evoke different conversational goals or speaker agendas.

Abbreviations

BDOTS = Bootstrapped Differences of Timeseries

Data accessibility statement

Experiment materials, as well as data analysis files, are available via https://osf.io/95fwh/.

Ethics and consent

Recording of audio stimuli was approved by the University of Edinburgh Ethics Committee with the reference number 119/2021-2. Experiment 1 was approved by the University of Edinburgh Ethics Committee with the reference number 84/2021-6, and Experiment 2 was approved with the reference number 84/2021-8.

Competing interests

The authors have no competing interests to declare.

Author contributions

Wei Li: Conceptualization; Data Curation; Formal Analysis; Investigation; Methodology; Visualization; Writing – Original Draft Preparation; Writing – Review & Editing.

Hannah Rohde: Conceptualization; Data Curation; Methodology; Visualization; Writing – Review & Editing.

Martin Corley: Conceptualization; Data Curation; Formal Analysis; Methodology; Visualization; Writing – Review & Editing.

References

Audacity (Version 2.4.2) [Computer software]. (2020). Retrieved from https://www.fosshub.com/Audacity-old.html?dwl=audacity-macos-2.4.2.dmg

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence: Resources for processing natural language* (pp. 241–301). Springer. https://doi.org/10.1007/978-94-009-2727-8_10

Beltrama, A., & Schwarz, F. (2024). (Im)precise personae: The effect of socio-indexical information on semantic interpretation. *Language in Society,* 1–8. https://doi.org/10.1017/S0047404524000320

Bonnefon, J. F., Dahl, E., & Holtgraves, T. M. (2015). Some but not all dispreferred turn markers help to interpret scalar terms in polite contexts. *Thinking & Reasoning, 21*(2), 230–249. https://doi.org/10.1080/13546783.2014.965746

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440. https://doi.org/10.1016/j.cognition.2012.11.012

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 696–702. https://doi.org/10.1037/0278-7393.34.3.696

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668. https://doi.org/10.1016/j.cognition.2006.10.010

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28. https://doi.org/10.3758/BF03195938

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*, 1–55. https://doi.org/10.3765/sp.8.11

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, *9*(1), 519–540. https://doi.org/10.1146/annurev-linguistics-031220-010811

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710. https://doi.org/10.1111/cogs.12171

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., & Green, C. R. (1982). Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology*, *3*(4), 291–312. https://doi.org/10.1207/s15324834basp0304_6

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, vol. 3, Speech acts* (pp. 41–58). Academic Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). MIT Press. https://doi.org/10.7551/mitpress/8957.003.0013

Hsu, N. S., Kuchinsky, S. E., & Novick, J. M. (2021). Direct impact of cognitive control on sentence processing and comprehension. *Language, Cognition and Neuroscience*, *36*(2), 211–239. https://doi.org/10.1080/23273798.2020.1836379

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology, 58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Khorsheed, A., & Gotzner, N. (2023). A closer look at the sources of variability in scalar implicature derivation: A review. *Frontiers in Communication, 8,* 1187970. https://doi.org/10.3389/fcomm.2023.1187970

Khorsheed, A., Price, J., & van Tiel, B. (2022). Sources of cognitive cost in scalar implicature processing: A review. *Frontiers in Communication, 7,* 990044. https://doi.org/10.3389/fcomm.2022.990044

Kissine, M., & De Brabanter, P. (2023). Pragmatic responses to under-informative *some*-statements are not scalar implicatures. *Cognition, 237,* 105463. https://doi.org/10.1016/j.cognition.2023.105463

Li, W., Rohde, H., & Corley, M. (2021). Veritable untruths: Autistic traits and the processing of deception. *Journal of Autism and Developmental Disorders, 52,* 4921–4930. https://doi.org/10.1007/s10803-021-05347-4

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science, 41*(S6), 1434–1456. https://doi.org/10.1111/cogs.12378

Loy, J. E., Rohde, H., & Corley, M. (2018). Cues to lying may be deceptive: Speaker and listener behaviour in an interactive game of deception. *Journal of Cognition, 1*(1), 42. https://doi.org/gq7nc2

Loy, J. E., Rohde, H., & Corley, M. (2019). Real-time social reasoning: The effect of disfluency on the meaning of *some*. *Journal of Cultural Cognitive Science, 3*(2), 159–173. https://doi.org/10.1007/s41809-019-00037-1

Nolte, C., Seedorff, M., Oleson, J., Brown, G., Cavanaugh, J., & McMurray, B. (2023). *bdots: Bootstrapped Differences of Time Series*. R package version 1.2.5, https://CRAN.R-project.org/package=bdots.

Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt? *Neuropsychologia, 51*(13), 2770–2780. https://doi.org/10.1016/j.neuropsychologia.2013.08.019

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical Methods in Medical Research, 26*(6), 2708–2725. https://doi.org/10.1177/0962280215607411

Politzer-Ahles, S., & Husband, E. M. (2018). Eye movement evidence for context-sensitive derivation of scalar inferences. *Collabra: Psychology, 4*(1), 3. https://doi.org/10.1525/collabra.100

R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Roberts, C. (1996). Information structure in discourse: Toward a unified theory of formal pragmatics. *Ohio State University Working Papers in Linguistics, 49,* 91–136.

Ronderos, C. R., & Noveck, I. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition, 238*, 105480. https://doi.org/10.1016/j.cognition.2023.105480

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language, 102*, 55–67. https://doi.org/10.1016/j.jml.2018.05.004

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association, 31*(1), 153–169. https://doi.org/10/bwwc6z

Simons, M. (2017). Local pragmatics in a Gricean framework. *Inquiry, 60*(5), 466–492. https://doi.org/10.1080/0020174X.2016.1246865

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Tomlinson Jr., J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*(1), 18–35. https://doi.org/10.1016/j.jml.2013.02.003

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind, 4*, 71–87. https://doi.org/10.1162/opmi_a_00035