This better be interesting: A speaker’s decision to speak cues listeners to expect informative content

Hannah Rohde\textsuperscript{1}, Jet Hoek\textsuperscript{2}, Maayan Keshev\textsuperscript{3},
and Michael Franke\textsuperscript{4}

\textsuperscript{1}Department of Linguistics & English Language, University of Edinburgh, Edinburgh, UK
\textsuperscript{2}Department of Language & Communication, Radboud University, Nijmegen, The Netherlands
\textsuperscript{3}Department of Linguistics, Tel Aviv University, Tel Aviv, Israel
\textsuperscript{4}Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Keywords: language processing; pragmatics; predictability; informativity; real-world plausibility

Abstract

In anticipating upcoming content, comprehenders are known to rely on real-world knowledge. This knowledge can be deployed directly in favor of upcoming content about typical situations (implying a transparent mapping between the world and what speakers say about the world). Such knowledge can also be used to estimate the likelihood of speech, whereby atypical situations are the ones newsworthy enough to merit reporting (i.e. a non-transparent mapping in which improbable situations yield likely utterances). We report three forced-choice studies (two pre-registered) testing this distinction between situation knowledge and speech production likelihood. Comprehenders are shown to anticipate situation-atypical meaning more when guessing content that a speaker announces (rather than thinks), that is said out of the blue (rather than produced when prompted), and that is addressed to a large audience (rather than a single listener). The results demonstrate that encountering speech—the recognition that someone has decided to say something instead of nothing—is sufficient to create a comprehension bias in favor of content that deviates from one’s real-world knowledge about typical situations.

Corresponding author: Hannah Rohde, Hannah.Rohde@ed.ac.uk

Copyright ©2021 The Author(s). Published by MIT Press. All Rights Reserved.
INTRODUCTION

The process of producing natural language requires making a number of informational decisions, both about what content to express and how much detail to include. These decisions reflect well-studied pressures related to efficiency and expressivity (e.g., Degen, Hawkins, Graf, Kreiss, & Goodman, 2020; M. C. Frank & Goodman, 2012; Franke & Jäger, 2016; Grice, 1975; Levy & Jaeger, 2007; Rubio-Fernandez, 2016), but they have primarily been studied in contexts in which a speaker’s production is already underway (e.g., modifier inclusion/omission) rather than content selection when a speaker is deciding what to say. If speakers favor informative and newsworthy content, a concomitant comprehension bias ought to arise such that listeners come to expect content that is interesting. This paper asks how comprehenders apply their real-world knowledge to guess what a speaker will say next — specifically do they use their knowledge of what happens in ordinary situations to anticipate utterances about situations that are extra-ordinary?

To illustrate, consider the passages about housing prices in (1) and whether comprehenders have different expectations for a value that denotes what Sue thinks someone paid (something close to the average housing price?) versus what Sue believes would be newsworthy enough to merit telling (something more extreme than the average?).

(1) a. Sue lives in New York. She thinks that her new neighbors bought their apartment for $____
b. Sue lives in New York. She told me that her new neighbors bought their apartment for $____

If there is no distinction between what a speaker thinks and what they say out loud, then the completions for (1-a) and (1-b) ought to align. Indeed, current models of language comprehension portray a fairly direct mapping between the world and what speakers say about the world, insofar as sentences about situation-typical meanings are reported to be easier to process than situation-atypical meanings (e.g. Kutas & Hillyard, 1980). Such models do not deny a role for informativity, but by emphasizing a comprehension preference for typicality and plausibility, they in effect depict language as a transparent modality that speakers use to convey what they observe in the world. In contrast, the informativity-driven

---

Language users of course do many things with language aside from conveying newsworthy information, but the use of language as a channel for information transfer nonetheless represents a fundamental reason to communicate.
approach we take here makes explicit a distinction between the prior probability of a certain meaning and the (inversely related) likelihood of a speaker choosing to produce an utterance to convey that meaning.

**BACKGROUND**

An information-theoretic approach to processing is apparent in a number of models of speech production (Aylett & Turk, 2004; A. Frank & Jaeger, 2008; M. C. Frank & Goodman, 2012; Gahl, 2008; Hale, 2006; Jurafsky, Bell, Fosler-Lussier, Girand, & Raymond, 1998; Levy & Jaeger, 2007; Piantadosi, Tily, & Gibson, 2011; Zerkle, Rosa, & Arnold, 2017) but has received less attention for modelling comprehension (cf. Rohde, Futrell, & Lucas, 2021; Sedivy, 2003). In production, speakers are more likely to mention elements that are real-world atypical — e.g., object color (YELLOW vs. BLUE BANANAS, Engelhardt, Bailey, and Ferreira (2006); Engelhardt and Ferreira (2014); Rubio-Fernandez (2016); Sedivy (2003)), object material (CERAMIC vs. WOOL BOWLS, Mitchell, Reiter, and Van Deemter (2013)), or the instrument used for an action (STAB WITH A KNIFE vs. ICE PICK, Brown and Dell (1987); Grigoroglou and Papafragou (2016); Lockridge and Brennan (2002)). In Brown and Dell’s (1987) classic production study on content selection, they show that while a particular object (a knife) may be the (presumed) preferred instrument for stabbing, the mention of that typical instrument is dispreferred. Rather, it is only when a story contains an atypical stabbing (with an icepick) that speakers prefer to mention the instrument. If it is the case that listeners track these real-world priors (see (2)) and speech production likelihoods (see (3)), then these probabilities should be reflected in their comprehension biases — we don’t expect a speaker to have encountered an icepick stabbing (one hopes) or a blue banana or a woolen bowl, but we would expect them to mention it if they had.

(2) $p(\text{INSTRUMENT(STAB)}=\text{KNIFE}) > p(\text{INSTRUMENT(STAB)}=\text{ICE PICK})$

(3) $p(\text{mention instrument}|\text{ICE PICK}) > p(\text{mention instrument}|\text{KNIFE})$

The relationship between speakers’ productions and listeners’ interpretations in such contexts is well captured by models that are built on principles of rational communication (Maxims of cooperative conversation Grice (1975), Rational Speech Act model M. C. Frank and Goodman (2012), rational redundancy Degen et al. (2020), efficiency and pertinence Rubio-Fernandez (2016), game theory Benz,
Jäger, and van Rooij (2006); Franke (2009)). Such models are relevant to understanding speakers’ choice among available forms, as well as the comprehenders’ response when such forms are used more and less felicitously (see work on scalar implicatures, (Augurzky, Franke, and Ulrich (2019); Hunt III, Politzer-Ahles, Gibson, Minai, and Fiorentino (2013); Spychalska, Kontinen, and Werning (2016), particularly using EEG to test the interplay of prior and likelihood for scalars, Werning and Cosentino (2017); Werning, Unterhuber, and Wiedemann (2019)). However, these models have been primarily applied to contexts in which speech is already underway and a speaker must make decisions about what elements to include/exclude. More recently, such an approach has been extended to account for a speaker’s choice to speak at all (Rohde et al., 2021) and has been framed within a Bayesian approach to informativity. In this approach, the probability of a particular meaning is represented by the prior \( p(\text{meaning}) \). The conditional probability of a speaker mentioning that situation given that it has occurred is captured by the likelihood \( p(\text{form}|\text{meaning}) \). These in turn combine to yield the overall probability of a particular form arising, summed over all the possible meanings that might be expressed with that form.

\[
(4) \quad p(\text{form}_i) \propto \sum_{\text{meaning}} p(\text{meaning}) \ast p(\text{utterance} = \text{form}_i|\text{meaning})
\]

There are several key insights afforded by the formulation in (4). First is that the prior and likelihood can each be considered in their own right — when a comprehender estimates the probability of encountering different utterance forms, their assessment reflects not only an estimate of whether the meaning is probable but also their estimate of whether a speaker would have selected a particular surface form to convey that meaning. Second is that the available surface forms can include silence — a comprehender’s calculation can include \( p(\emptyset|\text{meaning}) \). Indeed a comprehender should be surprised (and seek out alternative intended meanings) if a speaker formulates an utterance about content that is too easily inferable (see Kravtchenko & Demberg, 2015). Lastly, estimates of the prior and likelihood can be adjusted independently. The prior may shift if the context moves from the familiar real world to an alternative reality (e.g. Troyer & Kutas, 2018); the likelihood may adjust in more subtle ways depending on factors like who the speaker is, why they are speaking, or who they are speaking to. The studies presented here test this model, specifically the distinction between prior and likelihood (Experiment 1) and the possibility of systematic adjustments to the likelihood (Experiments 2 and 3).
Given that a speaker typically can opt to say nothing, the choice to formulate a non-null utterance presumably reflects the newsworthiness and relevance of the content (Sperber and Wilson (1995, 2004); for related work on discourse coherence, see Asher and Lascarides (2003); Hobbs (1985); Kehler (2002); Mann and Thompson (1988)), which in turn should guide comprehenders’ expectations about what they will hear. Relevance-driven inferences influence comprehenders’ moment-by-moment processing of upcoming content (e.g. ERP Xiang and Kuperberg (2015) and eye-tracking while reading Hoek, Rohde, Evers-Vermeul, and Sanders (2021)). In addition, comprehenders can be shown to favor messages that are sufficiently newsworthy to merit sending (faster reading times for a newsworthy message about socks that cost $100 than socks that cost $2; Rohde et al. (2021)). Such a finding is surprising for models that link situation typicality directly to processing ease (Bicknell, Elman, Hare, McRae, & Kutas, 2010; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kuperberg, 2021; Kutas & Hillyard, 1980; Matsuki et al., 2011; Stanovich & West, 1979). However, it does make sense for a model like (4), in which the assessment of utterance probability can incorporate a role for informativity via the likelihood term. While Rohde et al.’s reading-time results establish slower processing for situation-typical meanings compared with situation-atypical meanings, their studies do not probe the content of participants’ expectations — which meanings do comprehenders believe speakers are likely to have encountered in the world (the prior) versus have chosen to say about the world (the likelihood) and what factors affect these expectations?

The studies we present here use forced-choice tasks to test comprehenders’ guesses about an upcoming numeric value in a proposition. Experiment 1 manipulates the status of the proposition as either an individual’s internal thought versus an articulated utterance. Experiments 2 and 3 manipulate the context of production — a statement produced when prompted versus out of the blue and when addressed to a single listener versus a crowd. The results suggest that comprehenders estimate the likelihood of utterance production in favor of content that deviates from their own real-world priors and they do so in context-sensitive ways.

A natural question is how such results stack up against evidence that comprehenders can use context to ease their processing of real-world atypical content (e.g., peanuts in love, Nieuwland & Van Berkum, 2006). However, such evidence comes from passages in which real-world implausibility is made plausible (e.g., via a description of an anthropomorphized peanut singing a love song). Such work thus demonstrates the malleability of the prior but not a comprehension bias in favor of situation-atypical content.
EXPERIMENT 1: PRIOR VERSUS LIKELIHOOD

This first experiment tests comprehenders’ expectations about upcoming content when it constitutes a character’s reported thought versus their reported speech, see (5).

(5) Sarah is a woman from the US. Sarah has an acquaintance, Eric.

a. Eric thinks that Sarah went to ... restaurants last year.

b. Eric announced to me that Sarah went to ... restaurants last year.

We manipulate whether a character is said to THINK or ANNOUNCE something. Participants chose between two numeric values: a ‘low’ one to approximate the mean and a ‘high’ one that is expected to be more newsworthy. If participants expect speakers to transparently map thoughts into speech, then a character’s reported thoughts ought to parallel that character’s reported speech. If, however, participants distinguish between the prior probability of a situation occurring and the likelihood that a speaker would choose to produce a sentence about that situation, the THINK condition ought to yield estimates that are closer to participants’ real-world priors than the ANNOUNCE condition.

Method

Materials  Each of the 12 critical passages introduced an individual (Judith in (5)) and someone who would know that individual reasonably well (brother, Bill). The final sentence described this second person’s thought or announcement about some aspect of the first individual’s life (see Appendix A). The manipulation here and in Experiments 2 and 3 was implemented as a within-participants and within-items design. The two numeric values for each passage were selected via a pre-test (N=31, $2 Amazon Mechanical Turk task) where participants provided free responses to questions about the number of items or frequency of events in someone’s life (Sarah is a woman from the US. How many restaurants did Sarah go to last year?). The ‘low’ value was selected as a value slightly above that item’s pre-test mean (mean + 1/5*standard deviation) and the ‘high’ one as a value farther above the mean (mean + 4/5*standard deviation). Both values were ‘plausible’ in that they represented values in the range elicited in the pre-test, but the high values were less probable (and therefore more newsworthy). We rounded values to the nearest whole
number, except if one of the values for a specific item was a multiple of 5. In that case, we rounded one
of the values to the closest other whole number to make sure that both values were either a multiple of 5
or not in order to avoid a confound where a precise (non-multiple-of-5) value seemed more informative.

Participants also saw 8 filler passages (see Appendix D). Four fillers required speculation (e.g., *Amtrak
operates trains in the US. The passengers know that delays are common. Indeed, yesterday’s 08:30
Amtrak train from NYC to Boston was 27 / 37 minutes late.*), while four were catch trials with a correct
answer (e.g., *You have to stand on your own 2 / 6 feet.*). Participants who made mistakes on any catch
trials were excluded from analysis.

**Participants** 97 native-English speakers were recruited through Amazon Mechanical Turk and paid for
their participation ($2). We removed data from 7 participants who made mistakes on the catch trials,
leaving data from 90 participants (mean age 41.1, range 23-77; 31 women, 2 participants with no gender
specified).

**Data analysis** For all three experiments, we analyzed the binary outcome of participants’ forced-choice
selection (low versus high value) with logistic mixed effects models (GLMM: *Jaeger* (2008)) using the
lme4 package (*Bates, Mächler, Bolker, & Walker*, 2015) in R (*R Core Team*, 2019) with the maximal
random effects structure permitted by the data (*Barr, Levy, Scheepers, & Tily*, 2013). The categorical
fixed effect of *condition* was deviation-coded and its significance was determined by performing a
likelihood ratio test to compare the fit of the model to one with the same random effects structure but no
fixed effect.

**Results**

As predicted, the **ANNOUNCE** condition yielded more selections of the higher value than the **THINK**
condition ($\beta = 0.40, SE = 0.15, z = 2.66, p < .001$). Figure 1 shows a preference for the lower, more
typical, value in the **THINK** condition and a 50-50 split between the lower and higher values in the
**ANNOUNCE** condition.

**Discussion**
As predicted by a model in which expectations for newsworthiness influence comprehenders’ guesses about upcoming content, comprehenders showed a stronger preference for the situation-typical value (close to the estimated real-world mean) when the passage reported someone’s thoughts rather than their speech. The finding that the THINK condition does not strongly favor the lower value could reflect participants’ low sensitivity to the contrast between the chosen numbers, or their belief that the THINK sentences were themselves speech productions from a narrator (the experimenter) and thus may contain information that is (minimally) interesting enough to utter.

**EXPERIMENT 2: LIKELIHOOD OF SPEECH**

If comprehenders estimate utterance likelihood when making guesses about upcoming content, a question is whether that likelihood is malleable. If it is, certain discourse contexts may increase the expectation for newsworthiness —for example, spontaneous speech would be predicted to contain more newsworthy content than speech that is produced as an answer to a question.\(^3\)

---

\(^3\)This experiment was preregistered: osf.io/dhm5g
**Method**

**Materials** 35 experimental passages followed the structure from Experiment 1, except that the final sentence varied whether a character said something **OUT OF THE BLUE** or **WHEN ASKED** (Appendix B).

(6) Jeff is a man from the US. Jeff lives across the street from Amy. A few minutes ago,

a. Amy, **when asked about it**, said that Jeff spent . . . hours in his car last week.

b. Amy **out of the blue** said that Jeff spent . . . hours in his car last week.

As in Experiment 1, the values were selected via a free-prompt pre-test (N = 20, $4 Mturk task). Here, the lower value corresponds to the mean of the pre-test responses and the higher value to the mean + 1SD of the pre-test responses. The same 8 fillers were used from Experiment 1.

**Participants** 110 native speakers of English were recruited through Amazon Mechanical Turk and paid for their participation ($5). We removed 7 participants who made mistakes on the catch trials, leaving 103 participants (mean age 37.7, range 19-68; 42 women, 2 participants with no gender specified).

**Results**

As predicted, the **OUT OF THE BLUE** condition yielded more selections of the higher value than the **WHEN ASKED** condition ($\beta = -0.34, SE = 0.11, z = -3.16, p < .01$). Figure 2 shows a preference for the lower, more typical, value in the **WHEN ASKED** condition and a 50-50 split between the lower and higher values in the **OUT OF THE BLUE** condition.

**Discussion**

Experiment 2 shows that comprehenders prefer the atypical (newsworthy) value more when reported speech is spontaneous. This finding is again in line with the model put forward in (4). While participants’ baseline prior is unlikely to be affected by our manipulations, our results show that the discourse context informs participants’ estimate of the speaker’s content, presumably via the likelihood. The fact that the **WHEN ASKED** condition showed a substantial rate of higher value responses could, in addition to the
reasons mentioned in the context of Experiment 1, also be due to the question itself presupposing at least some potential newsworthiness of the value.

EXPERIMENT 3: AUDIENCE SIZE

This final experiment tests whether comprehenders use information about the speaker’s audience to adjust their expectations about upcoming content. If comprehenders can adjust their estimate of the likelihood of speech depending on both the typicality of the content and the size of the audience, we predict that the larger the audience, the more newsworthy the expected content will be.⁴

Method

Materials The 35 critical passages were adapted from Experiment 2 such that the reported speech was said TO ME or TO EVERYONE (see Appendix C).

⁴This experiment was preregistered: osf.io/6t5ze
(7) Jeff is a man from the US. Jeff lives across the street from Amy. A few minutes ago at a work
dinner,

a. Amy said to me that Jeff spent . . . hours in his car last week.

b. Amy stood up and said to everyone that Jeff spent . . . hours in his car last week.

The numeric values used were the same as in Experiment 2, as were the 8 filler items.

Participants 203 native speakers of English were recruited through Amazon Mechanical Turk and paid
for their participation ($5). We removed data from 51 participants who made mistakes on the catch trials,
leaving 152 participants (mean age 37.2, range 22-71; 56 women, 5 participants with no gender
specified).

Results As predicted, participants selected the higher value more in the TO EVERYONE condition than in the TO
ME condition ($\beta = 0.17, SE = 0.06, z = 2.59, p < .05$). As can be seen in Figure 3, the effect, though
statistically significant, is modest.

Discussion The results from Experiment 3 show that comprehenders expect the content of an utterance to be more
newsworthy when it is shared with a large group of people compared to when its audience consists of a
single person. Comparing Figure 3 to Figures 1 and 2 reveals that the proportion of high responses in the
TO ME condition in Experiment 3 closely matches the proportion of high responses in the ANNOUNCE
condition from Experiment 1 and the OUT OF THE BLUE condition from Experiment 2. This is to be
expected, since the prompts, though formulated slightly differently, correspond to similar conversational
scenarios: a speaker, of their own volition, decides to convey a piece of information in an utterance to a
(presumably) single other person.

GENERAL DISCUSSION AND CONCLUSION
Across three experiments, we measured comprehenders’ informativity expectations. Comprehenders filled in the blank with an atypical (high) value more in conditions where speakers announce something out loud (rather than think it), where they speak out of the blue (rather than when asked), and when they address a large audience (rather than a single listener). The act of choosing to convey content in speech, as well as the context of that speech, affects comprehenders’ expectations. These findings can be captured in a Bayesian approach in which the probability comprehenders assign to a particular utterance rationally combines the probability of the described situation ($p(\text{meaning})$) and the conditional probability that a speaker would articulate a linguistic form to describe such a situation to a certain audience ($p(\text{form}|\text{meaning})$). Our findings suggest that the prior and likelihood are separable and that the likelihood can be manipulated independently of the prior.

The first two experiments showed similar means in their respective low-informativity and high-informativity conditions. This is in line with our conjecture that question answering requires less “newsworthiness”. When the speaker is asked for a piece of information, the likelihood of speaking is high regardless of the content of the utterance. This leaves the prior as the crucial determinant of the
utterance probability. Therefore, the WHEN ASKED condition in Experiment 2 straightforwardly aligns with the THINK condition in Experiment 1.

In Experiment 3, we found that the portrayal of different audiences induces different guesses about a speaker’s upcoming content. This is in line with recent findings showing that manipulating the relationship between a speaker and addressee (stranger vs. family member) can alter comprehenders’ lexical predictions (Rubio-Fernandez, Mollica, Ali, & Gibson, 2019). Comprehenders show reading time slowdowns for utterances that violate assumptions about the common ground between speaker and addressee (“That greasy food is bad for your ulcer”, when spoken to a stranger), suggesting that comprehenders’ estimate of the likelihood of certain content varies depending on the speaker’s audience. Here, we likewise suggest that audience design has a role to play in modulating predictions, not through perceptions of interlocutor familiarity but through an expectation of speaker informativity.

It is worth noting that although the observed effects are statistically robust, the numeric differences seem fairly small. Overall selection rates in this study were close to chance level (ranging between 42-55%). The relatively small difference between conditions could be related to the fact that the two values that participants had to choose between were relatively similar. Only one standard deviation distinguishes the typical and atypical values, in all three experiments. Thus, it could be that participants are not fully aware of the contrast. It could even be that for some participants, the higher value is perceived as more probable, given that these higher values were named by some participants in the pretest. It is possible that with more prominently discriminated values, participants’ preferences would be even clearer.

Another possibility is that participants perceived the low-informativity conditions (THINK, WHEN ASKED, and TO ME) as still intended to be informative. Under this explanation, participants consider that there is a narrator, the experimenter, who reports the newsworthy thoughts and statements of different characters. Participants model not just the communicative goals of the characters in the passages but also the communicative goal of a hypothetical narrator—someone who is reporting these characters’ thoughts and statements because the thoughts and statements are newsworthy. A narrator could be informative by describing a character who thinks surprising thoughts or who boldly produces a highly uninformative utterance.
In Experiment 3, the contrast between the conditions was numerically smaller, relative to Experiments 1-2. The mean of the small audience (one person) condition numerically aligned with the out-of-the-blue and announcement conditions of the two previous experiments, and the larger audience condition seems to push expectations a bit further towards the atypical value. This could mean that the choice to spontaneously produce an utterance (rather than remaining silent) has more influence on informativity expectations than audience design considerations. However, this is merely speculative at this point and would require additional research. It is possible that the narrator-as-informative-speaker effect was enhanced in Experiment 3 because the passages invoked a first person narrator (“said to me”). The passages also described bolder communicative acts (i.e., speaking in front of a crowd), which perhaps are more likely to be retold by a narrator compared to, for example, someone’s thoughts in Experiment 1. Additionally, that there were not more choices of the ‘high’ value in Experiment 1 could also be due to a ceiling effect given the upper-bounding nature of prior plausibility.

To conclude, we argue that comprehenders consider both content plausibility and utterance likelihood, such that a ‘good’ utterance is one that balances the prior probability of the content with its novelty. Our focus on content selection goes beyond prior studies of rational speaker-listener behavior, by considering message-level production choices rather than the inclusion/omission of linguistic elements once an utterance is already underway. In addition, we find context-driven effects on comprehenders’ estimates of utterance likelihood. The current study thus emphasizes the importance of including a bias for informativity in models of language comprehension, a bias that may pull linguistic expectations away from situation-typical content. Importantly, this bias is not a uniform one but varies systematically with the speaker’s context of use. This sets the stage for additional psycholinguistic research to consider different metrics of what makes language use efficient and relevant.

REFERENCES


meeting of the cognitive science society (p. 3062-3067).

*Journal of Cognitive Neuroscience, 18*, 1098-1111.


APPENDIX A: ITEMS EXPERIMENT 1

1. Joseph is a man from the US. Joseph lives next door to Sue. Sue thinks / announced to me that Joseph ate . . . burgers last month. 8 / 11

2. Nick is a man from the US. Nick went to school with Stephanie. Stephanie thinks / announced to me that Nick saw . . . movies last year. 22 / 36

3. Melanie is a woman from the US. Melanie has a colleague, Bob. Bob thinks / announced to me that Melanie owns . . . pairs of shoes. 73 / 152

4. Erin is a first grade student in primary school. Erin has an uncle, Josh. Josh thinks / announced to me that there are . . . children in Erin’s class. 24 / 27

5. Betty is a woman from the US. Betty works at an office with David. David thinks / announced to me that Betty washed her hair . . . times last month. 21 / 27

6. Lelia is a woman from the US. Lelia lives around the corner from Brandon. Brandon thinks / announced to me that Lelia has . . . friends. 10 / 14

7. Tony is a man from the US. Nick has a sister, Emily. Emily thinks / announced to me that Tony cooked . . . meals at home last month. 12 / 17

8. Liam is a man from the US. Liam lives down the street from Rebecca. Rebecca thinks / announced to me that Liam has . . . T-shirts. 21 / 29

9. Judith is a woman from the US. Judith has a brother, Bill. Bill thinks / announced to me that Judith has . . . Facebook friends. 207 / 268

10. Andy is a man from the US. Andy has an aunt, Katherine. Katherine thinks / announced to me that Andy drank . . . cups of coffee last week. 14 / 20

11. Lisa is a woman from the US. Lisa has a friend, Kevin. Kevin thinks / announced to me that Lisa made . . . phone calls last week. 22 / 32
12. Sarah is a woman from the US. Sarah has an acquaintance, Eric. Eric thinks / announced to me that Sarah went to ... restaurants last year. 46 / 78

APPENDIX B: ITEMS EXPERIMENT 2

1. Joseph is a man from the US. Joseph has a girlfriend, Sue. Yesterday, when asked about it, Sue said that / Sue out of the blue said that Joseph ate ... burgers last month. 6 / 9

2. Nick is a man from the US. Nick went to school with Stephanie. Yesterday, when asked about it, Stephanie said that / Stephanie out of the blue said that Nick saw ... movies last year. 19 / 31

3. Melanie is a woman from the US. Melanie has a colleague, Amber. This morning, when asked about it, Amber said that / Amber out of the blue said that Melanie owns ... pairs of shoes. 12 / 18

4. Erin is a first grade student in primary school. Erin has an uncle, Josh. This morning, when asked about it, Josh said that / Josh out of the blue said that there are ... children in Erin’s class. 23 / 44

5. Betty is a woman from the US. Betty works at an office with David. Tonight, when asked about it, David said that / David out of the blue said that Betty washed her hair ... times last month. 18 / 29

6. Lelia is a woman from the US. Lelia lives around the corner from Brad. Tonight, when asked about it, Brandon said that / Brandon out of the blue said that Lelia has ... friends. 9 / 13

7. Tony is a man from the US. Nick has a sister, Emily. This afternoon, when asked about it, Emily said that / Emily out of the blue said that Tony cooked ... meals at home last month. 36 / 101

8. Liam is a man from the US. Liam lives down the street from Rebecca. Last week, when asked about it, Rebecca said that / Rebecca out of the blue said that Liam has ... T-shirts. 13 / 21

9. Andy is a man from the US. Andy has an aunt, Katherine. This afternoon, when asked about it, Katherine said that / Katherine out of the blue said that Andy drinks ... cups of coffee per day. 2 / 3

10. Lisa is a woman from the US. Lisa has a stepmother, Mona. Today, when asked about it, Mona said that / Mona out of the blue said that Lisa made ... phone calls last week. 17 / 26

11. Sarah is a woman from the US. Sarah has an acquaintance, Eric. Last week, when asked about it, Eric said that / Eric out of the blue said that Sarah went to eat out in a restaurant ... times last year. 24 / 49

12. Hugh is a man from the US. Hugh has a neighbor, Jenn. Just now, when asked about it, Jenn said that / Jenn out of the blue said that Hugh walked his dog ... times last week. 9 / 13
13. Nathalie is a woman from the US. Nathalie often has dinner with Alice. A few days ago, when asked about it, Alice said that Nathalie read... books last year. 7 / 11
14. Zach is a man from the US. Zach has a brother, Jim. A few days ago, when asked about it, Jim said that Zach went to... concerts last year. 3 / 5
15. Jess is a woman from the US. Jess takes sewing classes with Anna. On Tuesday, when asked about it, Anna said that Jess has... cousins. 6 / 9
16. Gary is a man from the US. Gary has a cousin, Alexander. On Tuesday, when asked about it, Alexander said that Gary was on... flights last year. 2 / 4
17. Samantha is a 14-year-old girl from the US. Samantha has an aunt, Caroline. On Wednesday, when asked about it, Caroline said that Samantha attended... birthday parties last year. 5 / 8
18. Jennifer is a woman from the US. Jennifer has a sister, Holly. On Wednesday, when asked about it, Holly said that Jennifer worked out... times last month. 11 / 16
19. Jill is a woman from the US. Jill has a best friend, Kevin. A few hours ago, when asked about it, Kevin said that Jill spent... hours on her phone last week. 15 / 24
20. Ralph is a man from the US. Ralph has an uncle, Harry. A few hours ago, when asked about it, Harry said that Ralph ordered take-out... times last month. 7 / 11
21. Brendan is a man from the US. Brendan has a gym buddy, Ryan. Just now, Ryan when asked about it, Ryan said that Brendan has... keys on his keychain. 5 / 7
22. Scott is a 12-year-old from the US. Scott has a classmate, Matt. Today, when asked about it, Matt said that Scott ate... candy bars last week. 9 / 14
23. Kyle is a teenager from the US. Kyle has a friend, Wade. A few minutes ago, when asked about it, Wade said that Kyle had football practice... times last month. 10 / 15
24. Robert is a man from the US. Robert has a co-worker, Margaret. About a week ago, when asked about it, Margaret said that Robert ordered something online... times last year. 24 / 78
25. Wanda is a woman from the US. Wanda carpools with Adam. Last Saturday, when asked about it, Adam said that Wanda has... cards in her wallet. 6 / 8
26. Jeff is a man from the US. Jeff lives across the street from Amy. A few minutes ago, when asked about it, Amy said that Jeff spent . . . hours in his car last week. 11 / 16
27. Pauline is a woman from the US. Pauline is in a reading group with Jack. Last Saturday, when asked about it, Jack said that Pauline has . . . house plants. 4 / 7
28. Peter is a man from the US. Peter shares an apartment with Jeffrey. Two days ago, when asked about it, Jeffrey said that Peter washed . . . loads of laundry last month. 7 / 11
29. Lily is a woman from the US. Lily has a nephew, Bob. On Monday, when asked about it, Bob said that Lily ran her dishwasher . . . times last month. 18 / 27
30. Patricia is a woman from the US. Patricia lives next-door to Nora. On Monday, when asked about it, Nora said that Patricia vacuumed . . . times last month. 6 / 9
31. Linda is a woman from the US. Linda plays tennis with Beth. About a week ago, when asked about it, Beth said that Linda has . . . colleagues. 13 / 19
32. John is a man from the US. John lives in the same apartment building as Anton. About an hour ago, when asked about it, Anton said that John drank . . . beers last month. 15 / 25
33. Gilly is a woman from the US. Gilly has a roommate, Denise. About an hour ago, when asked about it, Denise said that Gilly drank . . . glasses of wine last month. 8 / 13
34. Rob and Wendy are a couple from the US. Rob and Wendy have a friend, Katie. The other day, Katie, when asked about it, said that Rob and Wendy own . . . chairs. 8 / 12
35. Trey and Tina are a couple from the US. Trey and Tina live next door to Paul. The other day, when asked about it, Paul said that Trey and Tina hired a babysitter . . . times last year. 24 / 52

APPENDIX C: ITEMS EXPERIMENT 3
1. Joseph is a man from the US. Joseph has a girlfriend, Sue. Yesterday at the local bar, Sue said to me / stood up and said to everyone that Joseph ate . . . burgers last month. 6 / 9
2. Nick is a man from the US. Nick went to school with Stephanie. Yesterday at my party, Stephanie said to me / stood up and said to everyone that Nick saw . . . movies last year. 19 / 31

3. Melanie is a woman from the US. Melanie has a colleague, Amber. This morning, at the subway, Amber said to me / stood up and said to everyone that Melanie owns . . . pairs of shoes. 12 / 18

4. Erin is a first grade student in primary school. Erin has an uncle, Josh. This morning at the office, Josh said to me / stood up and said to everyone that there are . . . children in Erin’s class. 23 / 44

5. Betty is a woman from the US. Betty works at an office with David. Tonight at the restaurant, David said to me / stood up and said to everyone that Betty washed her hair . . . times last month. 18 / 29

6. Lelia is a woman from the US. Lelia lives around the corner from Brad. Tonight at our family dinner, Brandon said to me / stood up and said to everyone that Lelia has . . . friends. 9 / 13

7. Tony is a man from the US. Nick has a sister, Emily. This afternoon at our pottery class, Emily said to me / stood up and said to everyone that Tony cooked . . . meals at home last month. 36 / 101

8. Liam is a man from the US. Liam lives down the street from Rebecca. Last week at the conference, Rebecca said to me / stood up and said to everyone that Liam has . . . T-shirts. 13 / 21

9. Andy is a man from the US. Andy has an aunt, Katherine. This afternoon at the museum, Katherine said to me / stood up and said to everyone that Andy drinks . . . cups of coffee per day. 2 / 3

10. Lisa is a woman from the US. Lisa has a stepmother, Mona. Today at the post office, Mona said to me / stood up and said to everyone that Lisa made . . . phone calls last week. 17 / 26

11. Sarah is a woman from the US. Sarah has an acquaintance, Eric. Last week at our choir rehearsal, Eric said to me / stood up and said to everyone that Sarah went to eat out in a restaurant . . . times last year. 24 / 49

12. Hugh is a man from the US. Hugh has a neighbor, Jenn. Just now at the store, Jenn said to me / stood up and said to everyone that Hugh walked his dog . . . times last week. 9 / 13

13. Nathalie is a woman from the US. Nathalie often has dinner with Alice. A few days ago at the bus, Alice said to me / stood up and said to everyone that Nathalie read . . . books last year. 7 / 11

14. Zach is a man from the US. Zach has a brother, Jim. A few days ago at work, Jim said to me / stood up and said to everyone that Zach went to . . . concerts last year. 3 / 5

15. Jess is a woman from the US. Jess takes sewing classes with Anna. On Tuesday at the market, Anna said to me / stood up and said to everyone that Jess has . . . cousins. 6 / 9
16. Gary is a man from the US. Gary has a cousin, Alexander. On Tuesday at the baseball game,
   Alexander said to me / stood up and said to everyone that Gary was on . . . flights last year. 2 / 4
17. Samantha is a 14-year-old girl from the US. Samantha has an aunt, Caroline. On Wednesday at the
   library, Caroline said to me / stood up and said to everyone that Samantha attended . . . birthday
   parties last year. 5 / 8
18. Jennifer is a woman from the US. Jennifer has a sister, Holly. On Wednesday at the cafe, Holly said
   to me / stood up and said to everyone that Jennifer worked out . . . times last month. 11 / 16
19. Jill is a woman from the US. Jill has a best friend, Kevin. A few hours ago at our work meeting,
   Kevin said to me / stood up and said to everyone that Jill spent . . . hours on her phone last week. 15 / 24
20. Ralph is a man from the US. Ralph has an uncle, Harry. A few hours ago at the fair, Harry said to
   me / stood up and said to everyone that Ralph ordered take-out . . . times last month. 7 / 11
21. Brendan is a man from the US. Brendan has a gym buddy, Ryan. Just now on the street, Ryan said to
   me / stood up and said to everyone that Brendan has . . . keys on his keychain. 5 / 7
22. Scott is a 12-year-old from the US. Scott has a classmate, Matt. Today at the playground, Matt said
   to me / stood up and said to everyone that Scott ate . . . candy bars last week. 9 / 14
23. Kyle is a teenager from the US. Kyle has a friend, Wade. A few minutes ago at school, Wade said to
   me / stood up and said to everyone that Kyle had football practice . . . times last month. 10 / 15
24. Robert is a man from the US. Robert has a co-worker, Margaret. About a week ago at the cafeteria,
   Margaret said to me / stood up and said to everyone that Robert ordered something online . . . times
   last year. 24 / 78
25. Wanda is a woman from the US. Wanda carpools with Adam. Last Saturday at our drama group,
   Adam said to me / stood up and said to everyone that Wanda has . . . cards in her wallet. 6 / 8
26. Jeff is a man from the US. Jeff lives across the street from Amy. A few minutes ago at a work dinner,
   Amy said to me / stood up and said to everyone that Jeff spent . . . hours in his car last week. 11 / 16
27. Pauline is a woman from the US. Pauline is in a reading group with Jack. Last Saturday the park,
   Jack said to me / stood up and said to everyone that Pauline has . . . house plants. 4 / 7
28. Peter is a man from the US. Peter shares an apartment with Jeffrey. Two days ago at our school reunion, Jeffrey said to me / stood up and said to everyone that Peter washed . . . loads of laundry last month. 7 / 11

29. Lily is a woman from the US. Lily has a nephew, Bob. On Monday at the train, Bob said to me / stood up and said to everyone that Lily ran her dishwasher . . . times last month. 18 / 27

30. Patricia is a woman from the US. Patricia lives next-door to Nora. On Monday at the square, Nora said to me / stood up and said to everyone that Patricia vacuumed . . . times last month. 6 / 9

31. Linda is a woman from the US. Linda plays tennis with Beth. About a week ago at the mall, Beth said to me / stood up and said to everyone that Linda has . . . colleagues 13 / 19

32. John is a man from the US. John lives in the same apartment building as Anton. About an hour ago at the beach, Anton said to me / stood up and said to everyone that John drank . . . beers last month. 15 / 25

33. Gilly is a woman from the US. Gilly has a roommate, Denise. About an hour ago at the movies, Denise said to me / stood up and said to everyone that Gilly drank . . . glasses of wine last month. 8 / 13

34. Rob and Wendy are a couple from the US. Rob and Wendy have a friend, Katie. The other day at the shop, Katie said to me / stood up and said to everyone that Rob and Wendy own . . . chairs. 8 / 12

35. Trey and Tina are a couple from the US. Trey and Tina live next door to Paul. The other day at the swimming pool, Paul said to me / stood up and said to everyone that Trey and Tina hired a babysitter . . . times last year. 24 / 52

APPENDIX D: FILLER ITEMS

correct1 No one knows exactly when the Roman alphabet was first invented. The letters correspond roughly to spoken sounds but not exactly. There are . . . letters in the modern alphabet. 26 / 70
correct2 Johnny and his brother are looking forward to the holidays. They’ve made a gingerbread house and decorated the tree. Johnny’s favorite Christmas carol is ‘The . . . days of Christmas’ 14 / 12
correct3 My best friend always reminds me to take a break. She quotes Roald Dahl that ’a little nonsense now and then, is cherished by the wisest men.’ It’s true that there are only . . . hours in a day. 24 / 8
Everyone eventually has to leave home and make their way in the world. You have to face earning a living and doing your own laundry. You have to stand on your own ... feet. 6 / 2

Corey and Charlotte are professional tuba players. They recently had a conversation about how old the tuba actually is. They concluded that the tuba is at most ... years old. 151 / 217

The Grongitts went to a barbecue party again last weekend. Mr Grongitts was very bored. He decided he would not visit another barbecue party for at least ... weeks. 4 / 6

Gina discovered a new band on the internet. It’s called 'This will destroy you' and she immediately purchased their debut album. As soon as she can afford it she intends to buy ... more of their albums. 3 / 4

Amtrak operates trains in the US. The passengers know that delays are common. Indeed, yesterday’s 8:30 Amtrak train from NYC to Boston was ... minutes late. 27 / 37