Mis-Heard Lyrics: an Ecologically-Valid Test of Noisy Channel Processing

Moshe Poliak (moshepol@mit.edu)
MIT

Hannah Kimura (hk102@wellesley.edu)
Wellesley College

Edward Gibson (egibson@mit.edu)
MIT

Abstract

Experiments in psycholinguistics allow us to test hypotheses and build theories. However, psycholinguistic experiments often suffer from low ecological validity, because participants are often required to perform an unusual task in the face of unusual materials. In the current experiment, we test the predictions of Noisy Channel Processing in a naturalistic task: identifying the lyrics of a song. We conducted an experiment where participants heard short excerpts from songs and then indicated which one out of four possible transcriptions they had heard. We found that the predictions of Noisy Channel Processing bear out: options with higher prior and likelihood were chosen more often by participants as the perceived song lyrics. Thus, Noisy Channel Processing is successful in explaining the everyday phenomenon of mis-heard song lyrics. More broadly, this suggests that Noisy Channel Processing captures everyday language processing, and that it is not dependent on unnatural experimental tasks and materials.

Keywords: psycholinguistics; language processing; noisy channel processing; music

Introduction

Experiments in psycholinguistics often rely on unusual tasks with unusual materials. Based on such experiments, we build theories of language processing. One such theory of language processing is the Noisy Channel theory, which posits that comprehenders interpret the utterance (be it speech, text, or sign) by merging it with their prior expectations about the meaning and the form of the utterance (Shannon, 1949; Levy, 2008; Gibson, Bergen, & Piantadosi, 2013). However, many experiments that have tested Noisy Channel Theory relied on presenting participants with materials that had either implausible meaning or rare structure (Gibson et al., 2013; Poppels & Levy, 2016; Gibson et al., 2017; Liu, Ryskin, Futrell, & Gibson, 2020; Chen, Nathaniel, Ryskin, & Gibson, 2023; Poliak, Ryskin, Braginsky, & Gibson, 2023; Poliak, Malik-Moraleda, & Gibson, 2024). For example, Gibson et al. (2013) presented participants with sentences like "The mother gave the candle the daughter" and asked participants yes/no comprehension questions like "Did the daughter receive anything?" Unusual materials like these raise questions about the generalizability of experiments that use them: What do participants think that they are being tested on? What strange behaviors do experimental demands elicit? What is the limit of generalizations that can be made from responses to yes/no comprehension questions in a long list of strange sentences? In the current study, we use mondegreens—mis-heard lyrics—as a tool to test Noisy Channel Processing in an ecologically valid way: listening to music and recovering the lyrics.

In her essay *The Death of Lady Mondegreen*, American author Sylvia Wright wrote of a childhood experience that is familiar to most, if not all (Wright, 1954). She recounts her surprise at learning that her memory of a verse from the Scottish ballad *The Bonnie Earl O' Moray* had been erroneous all her life. She had remembered that the verse's lyrics were

'Ye Highlands and ye Lawlands,
Oh where have you been?
They have slain the Earl o' Moray
And Lady Mondegreen.'

However, she was surprised to find out that the last line of the verse is not *"And Lady Mondegreen"* but *"And laid him on the green."* But why did Wright mis-hear these lyrics? In the current study, we propose an explanation using the Noisy Channel Processing framework for why song lyrics are often mis-heard, and then we test our proposed explanation experimentally.

Noisy Channel Processing

According to Noisy Channel Processing, the goal of language processing is to correctly understand what the speaker wanted to say (Shannon, 1949; Levy, 2008). The difficulty lies in that, often, the message that the speaker intended is often corrupted by *noise*. In this sense, noise is anything that corrupts the intended message, encompassing speaker disfluencies (or unusual pronunciation due to to singing), environmental sounds (like instrumental music, in addition to vocals), and listener noise (like lapses of attention or memory constraints). Therefore, the listener works with the perceived message (M_p) to recover what the intended message (M_i) was. How can the listener infer what was intended given a corrupted message? Noisy Channel proposes that the listener does that by merging the perceived message with prior expectations, in a process of Bayesian reasoning.

Formally, the listener seeks to find the intended message (M_i) that maximizes the probability of the intended message given the perceived message (M_p; see Equation 1). By Bayes' rule, this quantity is proportional to the probability of the perceived message given the probability of the intended message, also called the *likelihood* ($P(M_p|M_i)$) times the probability of the intended message ($P(M_i)$), also called the *prior*. When the listener perceives a message, especially one that

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

has a low prior probability, the listener will consider alternative messages that are similar. Eventually, if one message that is *not* the perceived message maximizes $P(M_i|M_p)$, that alternative message may be interpreted to be the intention of the speaker. Now let us consider the role of the prior and the likelihood in the context of mondegreens.

$$P(M_i|M_p) \propto P(M_p|M_i)P(M_i) \quad (1)$$

The Prior

When the perceived message has a low prior probability, the listener might consider alternative utterances that have a higher prior probability. For example, in the ballad, "*Lady Mondegreen*" may have a higher prior probability than "*Laid him on the green*." In Modern American English, which Wright spoke as a child, *green* is a color, not a synonym for grass. This makes the original line implausible (to lay the slain Earl on the color green?). Implausible utterances have low prior probability: they are unlikely to be intended. In contrast, it is entirely plausible that those who had slain the Scottish Earl, had also slain the Scottish Lady—whose name, to an American child without extensive knowledge of Scottish names, could very well have been "Mondegreen." This is one reason for why Wright might have heard her version of the lyrics, and not the original.

Plausible sentences have higher prior probability than implausible sentences. Gibson et al. (2013) presented participants with plausible sentences and implausible (violating animacy) sentences. For example, participants may have read a sentence like "The mother gave the candle to the daughter" (plausible) or "The mother gave the daughter to the candle" (implausible). With every such sentence, they asked a yes/no comprehension question, like "Did the daughter receive anything?" For the first sentence ("The mother gave the candle to the daughter"), the literal response would be "yes." For the second sentence ("The mother gave the daughter to the candle"), the literal response would be "no." Participants were found to respond literally nearly all the time when presented with a plausible sentence, but, for implausible sentences, participants often replied non-literally. The authors interpret this pattern to mean that, when faced with a sentence that has a low prior probability (e.g., an implausible sentence), they may consider alternative sentences that are plausible and are more likely to be intended a priori. In this case, when reading a sentence like "The mother gave the daughter to the candle," participants may reason that the word *to* was erroneously inserted into the sentence, and that the sentence that had been intended was "The mother gave the daughter the candle." And, according to this more plausible alternative, the correct response is "yes": the daughter did receive something. Further experiments in this and alternative publications have shown that the prior is sensitive to the degree of noise in the experiment, the reliability of the speaker (Gibson et al., 2017), and the probability of the utterance given a preceding context (Chen et al., 2023). Moreover, prior expectations about the form (not just the meaning) of the utterance have been shown to influence interpretation in similar ways to plausibility. That is, less frequent structures are more frequently interpreted non-literally (Ferreira, 2003; Poppels & Levy, 2016; Liu et al., 2020; Keshev & Meltzer-Asscher, 2021; Poliak et al., 2023, 2024).

The Likelihood

The likelihood, $P(M_i|M_p)$, penalizes potential intended messages for their distance from the perceived message. That is, the likelihood is the highest when the perceived and the intended messages are one, and it decreases as the potential intended message grows more dissimilar to the perceived message. For example, the potential intended message *Lady Mondegreen* is quite similar to *laid him on the green* (this can be measured, for example, using the Levenshtein distance between the phonetic transcription of the phrases; Levenshtein et al., 1966). Thus, the likelihood of the potential intended message *Lady Mondegreen* is quite high, albeit lower than that of the potential intended message *laid him on the green*. In turn, the likelihood of *Lady Mondegreen* is higher than that of a yet different interpretation of the lyrics, like *"laden mounds of green"*. This is because the latter is more dissimilar from the perceived message than the former. Therefore, according to the likelihood, *laid him on the green* is most likely to be the intended message according to the listener, followed by *Lady Mondegreen*, followed by *laden mounds of green*.

How the distance between the perceived and intended message is computed depends on the *noise model*. Previous work on the noise model has framed it in terms of Levenshtein distance: deletions, insertions, and/or exchanges that could change the intended utterance into the perceived utterance. Gibson et al. (2013) presented participants with sentences that varied by plausibility and construction. For example, an implausible sentence like "The mother gave the daughter to the candle" uses a prepositional-object construction. The closest plausible sentence would be a plausible sentence in the double-object construction, "The mother gave the daughter ( ) the candle." If the latter was the intended sentence, then, to produce the implausible prepositional-object sentence, the word "to" was erroneously inserted into the sentence. Alternatively, if the implausible sentence used a direct-object construction, like "The mother gave the candle the daughter," the most similar plausible intended message would be the prepositional-object-sentence "The mother gave the candle to the daughter." If the latter sentence was intended, then, to arrive at the implausible direct-object sentence, a deletion of the word *to* was involved. Gibson et al. (2013) showed that deletions were a more likely edit type than insertions: sentences that involved potential deletions were more likely to be interpreted non-literally. They also showed that, the more edits were required to reach the perceived utterance from the potential intended utterance, the more frequently those sentences were interpreted literally (signaling a lower likelihood of the edits in question). This pattern of findings has been replicated by several studied (Poppels & Levy, 2016; Chen et

al., 2023) and extended from whole-word edits to the edits of the final letter/morpheme of the verb (Poliak et al., 2023).

Transcribing Song Lyrics

In the current study we use the recognition of song lyrics as a naturalistic test of the predictions of the Noisy Channel framework. The task of lyric recognition is a particularly good fit for investigating the Noisy Channel framework because language production in the form of singing is noisier than recordings of speech that are specifically made for an experiment, and yet they have a concrete ground truth of the words that the singer intended to convey: the published written lyrics. Singing is more noisy than speaking for several reasons. Phonologically, relative to speech, voiced stops and nasals may be exchanged, vowels become more centralized, and vowel intelligibility decreases in high pitches (Smith & Scott, 1980; Benolken & Swanson, 1990; Hollien, Mendes-Schwartz, & Nielsen, 2000; Johnson, Huron, & Collister, 2014). Prosodically, the singing rhythm interferes with word stress, resulting in less intelligible pronunciation (Johnson et al., 2014). In terms of vocabulary, song lyrics often involve infrequent or archaic words (Johnson et al., 2014). As a result, about 25% of words in lyrics are mis-heard (Collister & Huron, 2008), with substantial variability across genres (Condit-Schultz & Huron, 2015). Just as it is difficult for humans, the task of lyric transcription for machines has also been known to be substantially more challenging than regular speech transcription (Gupta, Yılmaz, & Li, 2019). In sum, while the recognition of lyrics is a common and natural task, it is also highly error-prone, making it an apt case study for the predictions of the Noisy Channel Framework.

Current Study

We conducted an experiment to test the Noisy Channel explanation of mis-heard lyrics. In our experiment, participants listened to short excerpts from songs. After listening to each excerpt, they were presented with four possible transcriptions of the lyrics: the true transcription, as well as three mondegreens (similar-sounding, incorrect transcriptions) that were generated by the experimenters. Participants were asked to select the transcription that matched what they had heard in the preceding song excerpt. Since our main question involves how the priors and likelihood affect the recognition of lyrics, we chose to constrain the possibility space of possible transcriptions and opted for a forced-choice task and not for a transcription task. We made 2 predictions: (1) According to the likelihood term in the Noisy Channel equation (Equation 1), the more similar a transcription is to the true lyrics, the more likely participants are to indicate that it is the correct one. That is, holding other information constant, the correct transcription should be more likely to be chosen by participants than the mondegreens. (2) According to the prior term in the noisy channel equation (Equation 1), the higher the prior of a transcription relative to the other transcriptions that the participant sees, the more likely it is to be selected by the participant. If our predictions bear out, it will provide evidence that Noisy Channel Processing explains language processing in the real world, beyond the experimental setting.

Method

Materials

Audio Materials In the experiment, participants listened to 37 audio excerpts (32 critical items, 5 catch items). The 37 excerpts were several seconds long, selected from 37 distinct songs. We chose songs from a wide range of genres and avoided famous songs (all the songs that we selected had less than 500,000 plays on Spotify at the time of selection). All excerpts had at least 0.5 seconds of instrumental music preceding the vocals. After extracting the excerpts from the songs, we RMS-normalized them to be the same volume on average and modified each excerpt to have a fade-in of 0.3 seconds to avoid jarring participants with unexpected music.

Mondegreens Each audio track was paired with 4 transcriptions: one correct transcription and 3 mondegreens (incorrect transcriptions). For each of the 32 critical items, we created mondegreens that differed only a little from the true lyrics while still being grammatical (See Table 1 for a sample stimulus from the experiment). For the 5 catch items, we intentionally generated 3 highly dissimilar mondegreens as an attention check.

Participants

We recruited 50 participants through Prolific who identified as English-speaking monolingual Americans. We excluded one participant due to a technical error (several items appeared twice throughout the experiment). Of the remaining 49 participants, 46 chose the correct transcription on all catch trials, 2 missed one catch trial, and 1 participant missed two catch trials. We excluded the latter from the analyses, remaining with 48 participants, each with 32 observations for the critical items.

Procedure

The study requested that participants use headphones, which we verified at the beginning of the experiment using a headphone check (Woods, Siegel, Traer, & McDermott, 2017). Participants adjusted their volume and began the experiment. In each trial, participants heard the audio excerpt, and, once it finished, were displayed with the four possible transcriptions (on every trial, the position of the correct transcription was random). Participants were asked to click on the transcription that matched what they had heard in the audio excerpt. Then, participants were asked to indicate if they had heard the song before. Once they clicked on their response, the next trial was triggered. The order of trials was randomized for each participant.

Prior and Likelihood

Each transcription was associated with the two core quantities in Noisy Channel Processing: a prior and a likelihood. To quantify the prior probability of sentences, we obtained

Table 1: The correct lyrics and the three mondegreens of a sample item from the experiment.

Type	Lyrics
correct	They have slain the Earl of Moray / And laid him on the green
mondegreen	They have slain the Earl of Moray / And Lady Mondegreen
mondegreen	They have slain the Earl of Moray / And ladies mount the green
mondegreen	They have slain the Earl of Moray / And laden mounds of green

surprisal for every transcription from the bert-base-uncased language model (Devlin, Chang, Lee, & Toutanova, 2018), with the help of the minicons library (Misra, 2022) with left-to-right within-word masking (Kauf & Ivanova, 2023). We then computed the reciprocal ($\frac{1}{surprisal}$) to arrive at a quantity that is positively correlated with the prior probability (otherwise, the prior measure would be inversely correlated with the prior probability), and then we normalized this quantity for each item. In other words, we arrive at a quantity that is a prior probability: the sum of priors for all four transcriptions for one audio track is 1. Finally, following past work in psycholinguistics showing that language processing is particularly sensitive to the log of the probability, we log-transformed the prior for purposes of visualization and modeling (Shain, Blank, Fedorenko, Gibson, & Schuler, 2022).

To quantify the likelihood of sentences, we transformed each transcription into IPA using the transcription dictionary hosted at https://tophonetics.com, which is based on the Open Carnegie Mellon University Pronouncing Dictionary. Words that were out of the dictionary were transcribed to IPA manually. We then removed all spaces and computed the Levenshtein distance (Levenshtein et al., 1966) between each mondegreen and the correct transcription for each item, weighing deletions, insertions, and deletions equally. This resulted in a discrete distance measure, such that the correct transcription had distance of 0, and mondegreens had a value above 0 [1]. Like with the prior, we computed the reciprocal of the Levenshtein distance ($\frac{1}{1+distance}$)[2] and normalized it, resulting in a quantity that, for all four transcriptions of each item, sums up to 1. Following the same logic as with the prior probability, we log-transformed the likelihood probability.

Results

Manipulation Check

The experimental manipulation was successful. Participants selected the correct transcription 56.9% of the time, meaning that, although the correct transcription was selected more often than mondegreens, mondegreens were selected quite often. Moreover, there was no one item where all participants selected either the correct transcription or a mondegreen. We then investigated which songs were familiar to participants (recall that at the end of each trial, participants indicated whether they had heard the song before). Overall, participants have indicated having heard a song before in 11.6% of trials. The analyses described below were conducted on both the full dataset and a subset that included only songs that had not been heard before. The results turned out to be very similar and without any difference in inference. We therefore report results from the full dataset for brevity.

Data Preparation

To analyze and visualize the data, we used R (R Core Team, 2023) and tidyverse (Wickham et al., 2019). We summarized the data by counting how many participants chose each transcription for each item. This resulted in a dataframe with a row per transcription, i.e., 32 (items) * 4 (transcriptions per item) = 128 rows. The counts ranged from 0 to 46, and the sum of the counts for each item was 48 (since each participant saw each item once). We then turned the counts into proportions by dividing each observation by 48[3]. In the end, the variable s of interest were the proportion of times the transcription was selected [0,1] (the dependent variable), the log prior of the transcription, and the log likelihood of the transcription. Descriptively, the log prior was positively correlated (Pearson's r = 0.462) with the proportion of choices associated with each transcription (Figure 1). Similarly, the likelihood, too, was positively correlated (Pearson's r = 0.533) with the proportion of choices associated with each transcription (Figure 2).

Inference

We conducted a Bayesian regression using the brms package (Bürkner, 2021) to investigate the robustness and generalizability of our findings. We regressed the proportion of times any transcription was chosen on the likelihood and prior (we did not model the interaction between the two, because it is highly collinear with the likelihood[4]). This resulted in the

[1] There is one exception, which is an item where the true transcription involved the word "sun" and a mondegreen involved the word "son," which is a homonym. For that specific mondegreen, the distance was 0, too.

[2] We added 1 in the denominator because the reciprocal would not be defined for correct transcriptions, which have a distance of 0.

[3] Traditionally, multiple-choice data would be analyzed using a logistic regression that predicts the logit probability of the correct response. However, in the current study, the question is not whether participants succeed in choosing the correct transcription; rather, we ask whether we can predict which transcription is more likely to be chosen by participants, and whether a transcription is correct is implicit in its likelihood.

[4] While it is possible to decrease collinearity by transforming the variables further, for example, using a log transform and centering, we chose not to do so to maintain the simplicity and face validity of the predictors.

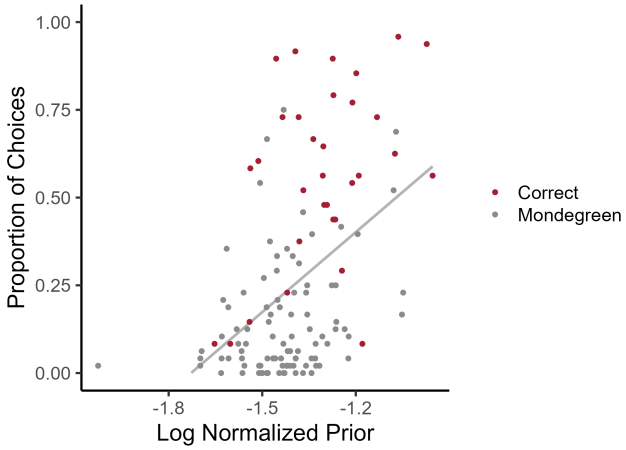

Figure 1: Transcriptions with higher log priors were selected more often by participants as the transcription that they had heard. Each point represents a transcription. Points in cardinal red represent the correct transcription, whereas points in silver gray represent mondegreens. The line of best fit represents the proportion of choice by participants as predicted by the log prior probability.

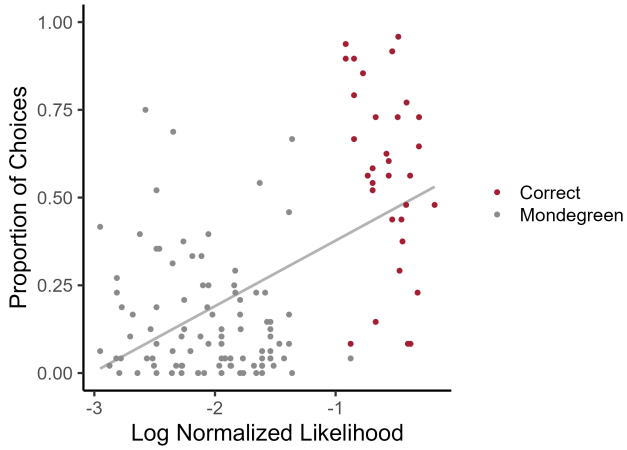

Figure 2: Transcriptions with higher log likelihoods were selected more often by participants as the transcription that they had heard. Each point represents a transcription. Points in cardinal red represent the correct transcription, whereas points in silver gray represent mondegreens. The line of best fit represents the proportion of choice by participants as predicted by the log likelihood.

formula $proportion \sim log\_prior + log\_likelihood$. We used 4 chains and 8000 iterations (4000 warmup). The model converged with no divergent transition and $\hat{R}$ values of 1 for all parameters. The model's estimate for the intercept was 1.26 (Median) with a 95% credible interval of [0.93, 1.59] and Estimated Error 0.16. The model found substantial evidence that transcriptions with higher log likelihoods were more likely to be selected by participants (Median = 0.15 95% Credible Interval = [0.10, 0.20], Estimated Error = 0.03) and that transcriptions with higher log priors were more likely to be selected by participants (Median = 0.54, 95% Credible Interval = [0.30, 0.79], Estimated Error = 0.12). Note the robustness of the effects, where for both likelihood and prior the lower bound of the credible intervals is more than 2 Estimated Errors away from 0.

The model that involves both log prior and log likelihood as predictors is substantially better than equivalent models with only one of the predictors. We compared the predictive ability of three types of models with the same specifications except for which predictors were included: (1) log likelihood and log prior (2) log likelihood only, and (3) log prior only. To evaluate the predictive ability of each model we used the Watanabe Akaike Information Criterion (WAIC) to compute the expected log pointwise predictive density (ELPD), which increases with more accurate prediction and decreases as the model becomes more complicated (Vehtari, Gelman, & Gabry, 2017). The comparison yielded only 0.8% estimates greater than 0.4, so we proceeded with the WAIC output. The model comparison identified the full model (prior + likelihood) as the one that predicts the data best (highest ELPD). The likelihood-only model was worse (ELPD difference = -8.3, standard error of difference = 4.4), and the prior-only model was the worst (ELPD difference = -14.2, standard error of difference = 5.2; see Figure 3).

Discussion

In the current study, we asked how experimental behavior in psycholinguistics generalizes to an ecologically valid task: recovering the lyrics of a song. When listening to songs we often do not hear the lyrics as they are, but some alternative, similar lyrics. For example, author Sylvia Wright reported hearing the words "Lady Mondegreen" when the correct lyrics were "laid him on the green" (Wright, 1954). We proposed that this phenomenon can be explained using the Noisy Channel Processing framework. According to Noisy Channel Processing, listeners actively try to infer the speaker's (or, in this case, the singer's) intended message given a perceived message. They do so by integrating the perceived message with their prior expectations. That is, when extracting words and morphemes from the perceived signal, listeners consider both "What words sound like what I just heard?" and "What words are likely to have been uttered?" The first question is quantified using the likelihood, or the distance between a certain transcription and the perceived message. The second question is quantified using the prior, or the expectations about the meaning and form of the utterance or lyrics. We quantified the likelihood using the Levenshtein distance between the true lyrics and each transcription. We quantified the prior probability using surprisal obtained from the large language model BERT. We found that, as predicted, the higher the likelihood and the prior probability of a transcription were, the more likely that transcription was to be se-

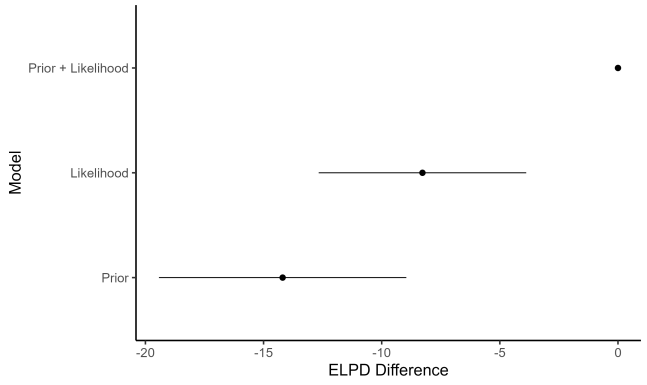

Figure 3: The expected log pointwise predictive density (ELPD) of each of the baseline models (likelihood-only or prior-only) relative to the main model (likelihood + prior), which had the best predictive strength. ELPD increases for better predictive ability and decreases for additional model complexity. Error bars are the ELPD difference standard error.

lected by participants as the perceived song lyrics. This suggests that the predictions of Noisy Channel Processing hold with natural tasks and materials, sidestepping issues of experimental demands and the effects of performing an unusual task while presented with a multitude of strange sentences in a questionnaire.

The effects of prior and likelihood that we found are strong and robust. First, the output of the Bayesian regressions showed that the lower bounds of the 95% credible intervals for prior and likelihood were more than 2 Estimated Errors away from 0, indicating robust effects. Second, we fit 2 baseline models with either only the prior or only the likelihood and compared them to the main model using WAIC, which found the full model to be the one the best in terms of its predictive ability when penalizing for its added complexity (Figure 3). This underscores the efficacy of Noisy Channel Processing in explaining language processing phenomena. Further work that investigates transcription in a noisy channel may benefit from prompting participants to transcribe speech, rather than choose between four transcriptions, which is even more naturalistic. In conclusion, in this study, we have applied a theoretical framework, Noisy Channel Processing, to an everyday phenomenon in language processing, mondegreens (mis-heard lyrics). By doing so, we have provided a scientific explanation for a common behavior and collected evidence that Noisy Channel Processing generalizes beyond the experimental setting and into day-to-day language processing.

References

Benolken, M. S., & Swanson, C. E. (1990). The effect of pitch-related changes on the perception of sung vowels. *The Journal of the Acoustical Society of America*, *87*(4), 1781–1785.

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. doi: 10.18637/jss.v100.i05

Chen, S., Nathaniel, S., Ryskin, R., & Gibson, E. (2023). The effect of context on noisy-channel sentence comprehension. *Cognition*, *238*, 105503.

Collister, L. B., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases.

Condit-Schultz, N., & Huron, D. (2015). Catching the lyrics: Intelligibility in twelve song genres. *Music Perception: An Interdisciplinary Journal*, *32*(5), 470–483.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, *abs/1810.04805*. Retrieved from `http://arxiv.org/abs/1810.04805`

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 164–203.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological science*, *28*(6), 703–712.

Gupta, C., Yılmaz, E., & Li, H. (2019). Acoustic modeling for automatic lyrics-to-audio alignment. *arXiv preprint arXiv:1906.10369*.

Hollien, H., Mendes-Schwartz, A. P., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice*, *14*(2), 287–298.

Johnson, R. B., Huron, D., & Collister, L. (2014). Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. *Empirical Musicology Review*, *9*(1), 2–20.

Kauf, C., & Ivanova, A. (2023, July). A better way to do masked language model scoring. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 925–935). Toronto, Canada: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2023.acl-short.80` doi: 10.18653/v1/2023.acl-short.80

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, 101359.

Levenshtein, V. I., et al. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In *Proceedings of*

the 2008 conference on empirical methods in natural language processing (pp. 234–243).

Liu, Y., Ryskin, R., Futrell, R., & Gibson, E. (2020). Structural frequency effects in noisy-channel comprehension. In *Proceedings of the 26th architectures and mechanisms for language processing.*

Misra, K. (2022). minicons: Enabling flexible behavioral and representational analyses of transformer language models. *arXiv preprint arXiv:2203.13112*.

Poliak, M., Malik-Moraleda, S., & Gibson, E. (2024). Sentence processing relies on expectations regarding both meaning and structure. In *2024 LSA annual convention.*

Poliak, M., Ryskin, R., Braginsky, M., & Gibson, E. (2023). It is not what you say but how you say it: Evidence from russian shows robust effects of the structural prior on noisy channel inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Poppels, T., & Levy, R. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. In *Cogsci.*

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience*, *42*(39), 7412–7430.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10–21.

Smith, L. A., & Scott, B. L. (1980). Increasing the intelligibility of sung vowels. *The Journal of the Acoustical Society of America*, *67*(5), 1795–1797.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*, 1413–1432.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi: 10.21105/joss.01686

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*, 2064–2072.

Wright, S. (1954). The death of lady mondegreen. *Harper's Magazine*, *209*(1254), 48–51.