Relating Scalar Inference and Alternative Activation: A View from the Rise-Fall-Rise Tune in American English

Thomas Sostarics, Eszter Ronai, and Jennifer Cole*

Abstract. The rise-fall-rise (RFR) tune in American English has received numerous theoretical accounts to describe its meaning contribution, with a consistent theme being the relationship between RFR and "higher alternatives." However, Autosegmental-Metrical theory predicts three RFR-shaped tunes which differ in the rising pitch accent used (H*, L+H*, L*+H), raising the question of whether different RFR-shaped tunes in fact behave differently. We investigate this question under the lens of scalar inference (SI). We find that RFR-shaped tunes with different pitch accents behave similarly in offline interpretation, increasing the rate of SI calculation relative to falling tunes. In online processing using cross-modal priming with lexical decision, we find an asymmetry in the processing profile of two RFR-shaped tunes: H*L-H% leads to additional facilitation of the higher alternative, while L*+HL-H% leads to less facilitation. We describe these results in relation to differences in pitch range and discuss how they relate to ongoing debates about RFR.

Keywords. intonation, prosody, scalar inference, priming, rise-fall-rise

1. Introduction. The Rise-Fall-Rise (RFR) tune in American English has received ample theoretical attention in semantics and pragmatics over the past forty years. The meaning contribution of RFR has been variably described as conveying uncertainty with regard to a higher value along a scale (Ward & Hirschberg 1985, Hirschberg & Ward 1992); conveying the existence of disputable higher alternatives (Constant 2012); highlighting the salience of a higher alternative (Göbel 2019, Göbel & Wagner 2023); indirectly or partially addressing a Question Under Discussion (QUD) (Wagner et al. 2013) or relating to a (hierarchically higher) contrastive topic or secondary QUD (Büring 2003, Westera 2019). Recent empirical work on RFR (de Marneffe & Tonhauser 2019, Buccola & Goodhue 2023, Ronai & Göbel to appear) has sought evidence for or against these accounts using experimental methods, finding variation in how adequately different theoretical proposals account for their results. While accounts of RFR will often describe a singular RFR, typically[1] referencing the (ToBI annotated) L*+HL-H% contour described by Ward & Hirschberg (1985), Autosegmental-Metrical (AM) phonological theory predicts not one but three putatively distinct RFR-shaped tunes which differ in the choice of rising pitch accent: H*, L+H*, or L*+H (Pierrehumbert 1980). Are the three RFR-shaped tunes interpreted similarly, indicating a broad class of RFR-shaped tunes, or do they behave differently from one another? We investigate potential contrasts between the RFR-shaped tunes under the lens of scalar inference (SI) in both offline

*We thank Kate Sandberg for help with recording, Michael Tabatowski for feedback on the written stimuli, and Chun Chan for experiment implementation support. We also thank the ProSD Lab and Experimental Meaning Group at Northwestern and the audiences at the Voices in Contexts workshop and Experiments in Linguistic Meaning for feedback on this work. Authors, all at Northwestern University: Thomas Sostarics (tsostarics@u.northwestern.edu) & Eszter Ronai (ronai@northwestern.edu), & Jennifer Cole (jennifer.cole1@northwestern.edu).

interpretation and online processing.

The conceptual link between RFR and SI is as follows. When describing the various pragmatic accounts for RFR, what stands out across accounts (rhetorically, though not necessarily implementationally) is a persistent invocation of some notion of "**higher** alternatives." SI is famously another domain within semantics/pragmatics where higher alternatives play a role. In the (Neo-)Gricean tradition, SI is taken to arise via reasoning about what a speaker could have said but did not (Grice 1975) with particular attention to pairs of lexical items that form a lexical scale (i.a., Horn 1972). Such scales are described in terms of relative informativity and can be formalized using a relation of asymmetric entailment (Horn 1972) such that for a scale $<X, Y>$, an utterance containing Y entails one containing X but not the other way around; hence, Y is the informationally stronger member of the pair and is often referred to as the *stronger* or *higher* scalemate while X is the *weaker* or *lower* scalemate. In (1), *some* is the weaker scalemate of the $<$*some*, *all*$>$ scale, leading the listener to reason about the speaker's use of *some* instead of *all*, ultimately arriving at the SI-enriched interpretation of *Jane ate some* ***but not all*** *of the cookies* from the literal meaning of the sentence, which is compatible with the alternative *Jane ate all of the cookies*.[2]

(1) Jane ate some of the cookies.
Jane ate some, and possibly all, of the cookies. LITERAL
Jane ate all of the cookies. ALTERNATIVE
$\rightsquigarrow$ Jane ate some but not all of the cookies. SI-ENRICHED

In the context of SI computation, some accounts of RFR make competing predictions regarding whether SI should be more or less likely when a sentence like (1) is uttered with RFR. Specifically, if the use of RFR conveys uncertainty about whether a higher alternative Y is true or not, then SI—the negation of Y—would be incompatible with such uncertainty. For instance, one cannot be uncertain whether Jane ate all the cookies while simultaneously concluding that Jane did not eat all the cookies. Recent experimental work on RFR has also relied on SI as its empirical testing ground, and has found that the use of RFR increases the likelihood of SI calculation (de Marneffe & Tonhauser 2019, Ronai & Göbel to appear, though cf. Buccola & Goodhue 2023).[3]These results thus appear prima facie as evidence against uncertainty accounts of RFR, but it remains an open question whether a different pattern of results, potentially in line with uncertainty accounts, might obtain for a different RFR-shaped tune.

[1]There is some variation in how researchers describe RFR. For example, Büring (2003; 537) describes an (L+)H*L-H% tune for contrastive topic marking, which Constant (2012; 431) argues is distinct from L*+HL-H%. Westera (2019; 326) notes a potential difference between H*L-H% and L*+HL-H%, but nonetheless elects to group the two together when comparing Büring and Constant's (among others') accounts of RFR to one another. Similarly, Wagner et al. (2013; 130) annotates RFR with an L+H* accent, attributing this to Hirschberg & Ward (1992) who specifically differentiate L*+H from L+H* in the context of RFR (*ibid* 242, see also Ward & Hirschberg 1985; 748). Regardless, the historical development of this literature has sought to identify the meaning contribution of "RFR," but it is not clear when, or whether, such accounts are expected to hold for putatively different RFR-shaped tunes.

[2]The likelihood that such SI-enriched interpretations arises varies across lexical scales, termed the *scalar diversity* phenomenon (Van Tiel et al. 2016). In work on scalar diversity, a primary goal is to understand the systematic factors (semantic, pragmatic, or otherwise) that contribute to this variation in rates of SI computation (Gotzner et al. 2018, Ronai & Xiang 2021, 2024, a.o.). The present work does not seek to explain scalar diversity, but rather uses it as a testing ground for competing predictions from existing accounts of RFR.

[3]While de Marneffe & Tonhauser (2019; 12) interpret their empirical results as showing that RFR "strengthens

In addition to probing the offline interpretation of different RFR-shaped tunes, one might also wonder whether there is a psycholinguistic processing correlate sensitive to the relationship between RFR and higher alternatives (and again, whether this varies for different RFR-shaped tunes). We test this question using cross-modal priming with lexical decision, which has previously been used to identify the activation of focus alternatives attributable to different intonational features (Braun & Tagliapietra 2010, Husband & Ferreira 2016, Gotzner et al. 2016). Specifically, contrastive focus alternatives (Rooth 1992) show sustained activation later in processing while words that cannot serve as contrastive focus alternatives, yet are nonetheless semantically related to the focus-accented element in a sentence, become deactivated later in processing. In our cross-modal lexical decision experiment, we extend prior text-only priming studies investigating scalar alternatives (Ronai & Xiang 2023, Lacina & Gotzner 2024) to not only investigate whether higher scalar alternatives are activated similarly to focus alternatives but, crucially, whether RFR has a modulating effect on this activation.

In summary, discussion of RFR often brings with it discussion of higher alternatives. Higher alternatives also play a key role in SI. Our goal is to use SI as a testing ground to identify potential differences between three putatively contrastive RFR-shaped tunes. We present experimental results from an inference judgment task, used to assess offline interpretation, and a cross-modal lexical decision task, used to assess online processing of RFR.

2. Materials and Norming. We use both cross-modal priming with lexical decision and the inference judgment task (henceforth just "inference task") using the same set of auditorily-presented materials. Because RFR cannot be used out of the blue, we need to have a preceding context before an utterance with RFR. Accordingly, we wrote indirect polar question-answer (Q/A) pairs between two speakers like in (2), where Bob's response uses either the lower alternative *cool* or the higher alternative *cold*.

(2) Alice: Did someone leave a window open in the office overnight?
Bob: The office feels *cool / cold*.

These Q/A pairs differ from those used in de Marneffe & Tonhauser (2019) because Alice's question does not use the higher alternative *cold*. This constraint on the materials is needed due to the lexical decision task: if we are interested in the activation status of *cold* following Bob's reply "The office feels cool.", then having *cold* explicitly mentioned in the preceding utterance (Alice's question) will directly activate it and likely mask any potential priming effect from intonation (see Gotzner et al. 2016 and Yan & Calhoun 2019 for effects of mentioned vs unmentioned alternatives in lexical decision). Lastly, these contexts are written to neither bias towards nor against SI calculation when Bob's answer contains the lower alternative. A literal interpretation of *The office feels cool (and possibly cold)* may be taken to be a positive response to the question (i.e., yes someone left a window open, explaining the chilliness) and an SI-enriched interpretation of *The office feels cool (but not cold)* may be taken to be a negative response to the question (i.e., no, nobody left a

the degree of belief in the scalar implicature," Ronai & Göbel (to appear; 447) interpret their findings, which also show an increase in SI computation, as RFR enhancing the salience of the higher alternative (Göbel 2019, Göbel & Wagner 2023), where increased salience of the higher alternative has been shown (through other means) to make SI more likely (Ronai & Xiang 2021). Other accounts may be compatible with either an increase or decrease in SI rates, e.g., Constant (2012), Westera (2019).

window open), but SI calculation itself is not **required** to arrive at a yes or no response (though, a relevance implicature is indeed needed).

In total, we wrote contexts for 74 adjectival scales reported in prior work (Aparicio & Ronai 2023, de Marneffe & Tonhauser 2019, Van Tiel et al. 2016) to use as our critical trial stimuli in both experiments. We also wrote an additional 122 filler trials, which comprise two types. First, 61 are Q/A pairs with non-word targets in the lexical decision task—we will refer to these as the **non-word fillers**. Second, to avoid a task adaptation effect where participants might learn that the only time they need to respond *Yes* is when they encounter a sentence-final adjective, we also include 61 Q/A pairs adapted from Husband & Ferreira (2016), example shown in (3), such that the accented word in the sentence (bolded) is placed at the end of Bob's response analogously to the critical trials—we will refer to these as the **HF16-adapted items**.

(3) *Original Item: The museum thrilled the **sculptor** when they called about his work.*
Alice: Did the museum deliver any good news?
Bob: The museum thrilled the **sculptor**.

Our critical materials were normed in a preliminary text-only task, which included naturalness rating and inference judgment components. Based on the rating results we removed 10 scales from our task that received a high proportion of low ratings, leaving a total of 64 critical items.

The remaining critical items were recorded by the first author in six intonation conditions: three RFR-shaped tunes (using L-H% edge tones) and three falling tunes (L-L%) that each differ in pitch accent (H*, L+H*, and L*+H). These recordings were then modeled using generalized additive models to create consistent targets for resynthesis (via PSOLA in Praat Boersma & Weenink 2020) for each tune in order to reduce the amount of variation across sentences. The averages of each tune across all the resynthesized recordings are shown in Figure 1 and are evenly distributed across all items.[4]

3. Inference Task. We recruited 84 participants from the online crowdsourcing platform Prolific for our inference task. This task is similar to prior work on SI calculation (Van Tiel et al. 2016, Ronai & Xiang 2021) where, on each trial, participants listened to a pre-recorded dialogue such as (2), where the answer includes the lower alternative from a lexical scale such as <*cool*, *cold*>. They were then presented with a question like *Would you conclude that the office does not feel cold?* and had to answer with "Yes" (indexing SI calculation) or "No". We included an additional 72 fillers, 36 of which come from the non-word fillers and 36 of which come from the HF16-adapted items. The items are distributed into 12 counterbalanced lists and then presented in a pseudorandom order that minimizes adjacent trials having the same intonational tune or item set. Each participant thus sees 136 trials (64 critical plus 72 fillers) divided into four blocks.

3.1. RESULTS. The average empirical SI rates for each intonation condition are shown in Figure 2. We can observe that the RFR-shaped tunes as a group show higher average SI rates than the falling tunes with potential graded distinctions between the three based on the pitch accent used.

We analyze these results with Bayesian logistic mixed effect regression models using brms (Bürkner 2021) with weakly informative regularizing priors. We include a fixed effect of tune, the

[4]While not shown here, our materials show previously described patterns of pitch accent peak height and alignment across accent types (Iskarous et al. 2024); an extended phonetic discussion must be left to a future paper.

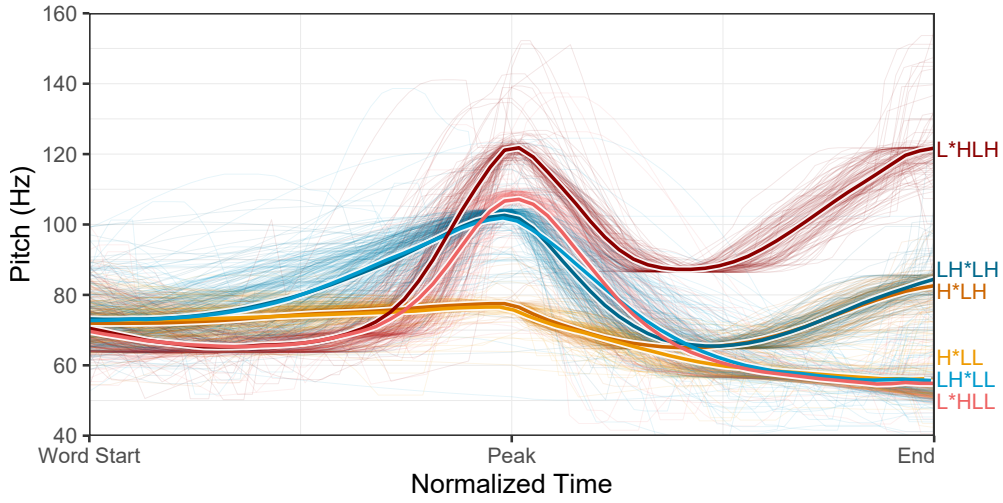

Figure 1: Resynthesized utterances for each tune, time-normalized to the location of the pitch accent peak. The averages of each tune are superimposed on top of the individual contours and labeled at the location of the final F0 value on the right-hand side.

combination of pitch accent (H*, LH*, L*H) and edge-tone configuration (LL, LH) and random intercepts by item and participants and random slopes of tune by participant and item. What we're interested in is the overall effect of edge-tone configuration (i.e., RFR-shaped versus falling) and the degree to which there are graded distinctions across the pitch accents, which we operationalize in terms of the differences of LH*−H* and L*H−H* **within each broad tune class** (RFR and falling). We use a manually-specified contrast matrix to encode these comparisons in a single model. Our analysis code is available at `https://osf.io/bc6a2/`.

From the statistical model we find a main effect of edge-tone configuration ($\hat{\beta} = 0.35, CI = [0.18, 0.52]$) such that the RFR-shaped tunes together yield higher SI rates than the falling tunes. The distinctions between pitch accents are small, and unsurprisingly given the overlap between the pitch accents evident in Figure 2, the 95% credible intervals for each pitch accent comparison contain 0. Although, the bulk of the posterior distribution for the L*HLH−H*LH comparison ($\hat{\beta} = 0.18, CI = [-0.08, 0.45]$) is greater than 0, with a probability of direction of 90.7%.

These results replicate prior work on RFR in the context of SI (de Marneffe & Tonhauser 2019, Ronai & Göbel to appear) but provide a novel finding that there is a primary dichotomy between broad falling and RFR-shaped tune classes. Moreover, this effect persists even when the higher alternative is not explicitly mentioned in the question (c.f. Ronai & Göbel to appear a.o.). While there appear to be numerical differences between pitch accents within these broad classes, there remains ample uncertainty as to the magnitude and systematicity of such distinctions. This caveat is perhaps not entirely surprising in light of work on variation in intonational form showing overlap in intonational categories in both production and perception (Arvaniti 2019, Cole et al. 2023, Steffman et al. 2024). Such overlap has also presented itself in psycholinguistic investigation, where despite much work drawing a convenient categorical distinction between "neutral" H* and "focus-marking" or "contrastive" L+H*, contrastive interpretations are nonetheless possible

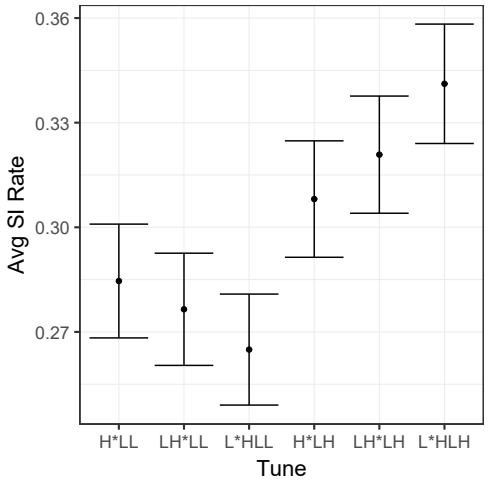

Figure 2: Average empirical SI rates for each tune with ± 2 standard errors.

even when H* is used (Watson et al. 2008). With regard to prior accounts of RFR, our results are compatible with theoretical proposals from Constant (2012), Göbel (2019), Westera (2019) but not accounts that require the truth of the higher alternatives to remain open through conveying uncertainty (Ward & Hirschberg 1985) or leaving the alternative unresolved or only partially addressed (Wagner et al. 2013).

4. Cross-modal Priming. Based on prior work on RFR, our primary hypothesis is that RFR evokes higher alternatives. In priming terms, we predict that when RFR is used with an utterance containing a lower alternative like *cool*, we will see facilitation in lexical retrieval of the corresponding higher alternative *cold* due to RFR boosting its activation level. Furthermore, since SI is uni-directional, we also expect that when RFR is used with a higher alternative, we will not see facilitation of the lower alternative; for instance, if RFR is used with *cold*, we might expect it to evoke a higher alternative like *freezing* but not a lower alternative like *cool*. Whether the lower alternative is specifically inhibited or simply not affected by RFR is left unspecified. Based on the results of Exp. 1, we expect that the three RFR-shaped tunes will behave similarly, predicting facilitation of lexical retrieval of the higher alternative for all three tunes, relative to falling tunes.

This task presents 186 trials split into six blocks of 31 trials. Of these, 64 trials are critical items which vary by intonational tune and by whether the higher or lower alternative serves as the visual target (with the other scalemate serving as the auditory prime). We will refer to the condition where the higher alternative is the target (i.e., hear *cool* then see *cold*) as the HIGHERTARGET condition and the condition where the lower alternative is the target (hear *cold* then see *cool*) as the LOWERTARGET condition. The fillers include 61 non-word fillers and 61 HF16-adapted items with the six intonational tunes evenly distributed across the items. Each HF16-adapted item has one of three possible targets; using an auditory prime of *sculptor* as an example, the three targets are a CONTRASTIVE semantic associate (i.e., a contrastive focus alternative) like *painter*, a NONCONTRASTIVE associate like *statue*, and a semantically UNRELATED word like *register*.

We recruited 104 undergraduate students from Northwestern University, who participated for

course credit. 41 participants were omitted due to language background or issues with completing the task, leaving 63 participants for analysis. On each trial, participants listened to a recorded audio dialogue. The visual target appeared on the screen 750ms after the offset of the final word of the utterance. Participants were instructed to use a button box to judge whether this visual target is a word or not a word of English (framed as YES, it is a word, or NO, it is not a word). The experiment was implemented in Psychopy (Peirce 2007) and administered in a sound-attenuated booth with a 165Hz monitor; participants gave their responses using a Cedrus RB-740 buttonbox.

4.1. RESULTS. Before we investigate the effect of intonation on reaction times, we first look at how our critical items compare to the HF16-adapted items. One condition in the HF16-adapted items is the CONTRASTIVE condition, where the target word can serve as a (contrastive) focus alternative to the prime word. Similarly, our adjectival scalemates, e.g., *cool* and *cold*, can **also** serve as focus alternatives to one another. However, our critical items additionally comprise a lexical scale and are hence related via asymmetric entailment. Accordingly, we are interested in first evaluating whether lexical scalemates offer any processing advantage **beyond** their status as focus alternatives. By doing so, we also gain a baseline for the RTs in each condition (i.e., what can be attributed to the relationship between the lexical items of the auditory prime and visual target) before seeing how intonation further modulates these.

We analyze participant reaction times when correctly responding YES (total accuracy 98.6%) using a Bayesian lognormal distributional model. Our main predictor of interest is target condition[5] controlled for effects of log word frequency (Balota et al. 2007) of the target word, length of the target word, and experimental block (all treated as continuous and centered). We include random intercepts by participant and item and random slopes of condition by participant; to account for differences in RT dispersion we also include random intercepts for the sigma parameter of the model by participant and item. The model predicted reaction times are shown in Figure 3.

From the statistical model, we find that RTs for higher alternatives are not credibly different from lower alternatives ($\hat{\beta} = 0, CI = [-0.03, 0.02]$). Moreover, the RTs for these scalar alternatives are not credibly different from contrastive alternatives ($\hat{\beta} = 0.01, CI = [-0.01, 0.03]$). RTs for noncontrastive associates are slower than contrastive and scalar alternatives ($\hat{\beta} = 0.05, CI = [0.02, 0.07]$) while semantically unrelated words are slower than all semantically related words ($\hat{\beta} = 0.09, CI = [0.06, 0.11]$)—this result shows the same pattern as the focus intonation condition in Husband & Ferreira (2016; 227). Overall, these results suggest that when the visual target can serve as a contrastive alternative to the auditory prime, there is not an additional processing advantage when the target and prime words are additionally related via asymmetric entailment.

Given that the two critical conditions show strong facilitation compared to the semantically unrelated filler condition, does intonation modulate the degree of facilitation within each condition? That is, does intonation contribute anything beyond the priming induced by the visual prime being a possible scalar alternative? Based on the hypothesis that RFR invokes higher alternatives and our offline results from Exp. 1 showing that the RFR-shaped tunes behave similarly to one another, we predict that the three RFR-shaped tunes will lead to additional facilitation in the HIGHERTARGET condition (i.e., *cool* uttered with RFR leads to faster lexical decisions of *cold*). Moreover,

[5]Condition has five levels and is helmert coded to encode nested orthogonal comparisons, which are depicted in Figure 3. Estimates are on the $\log_e$ RT scale.

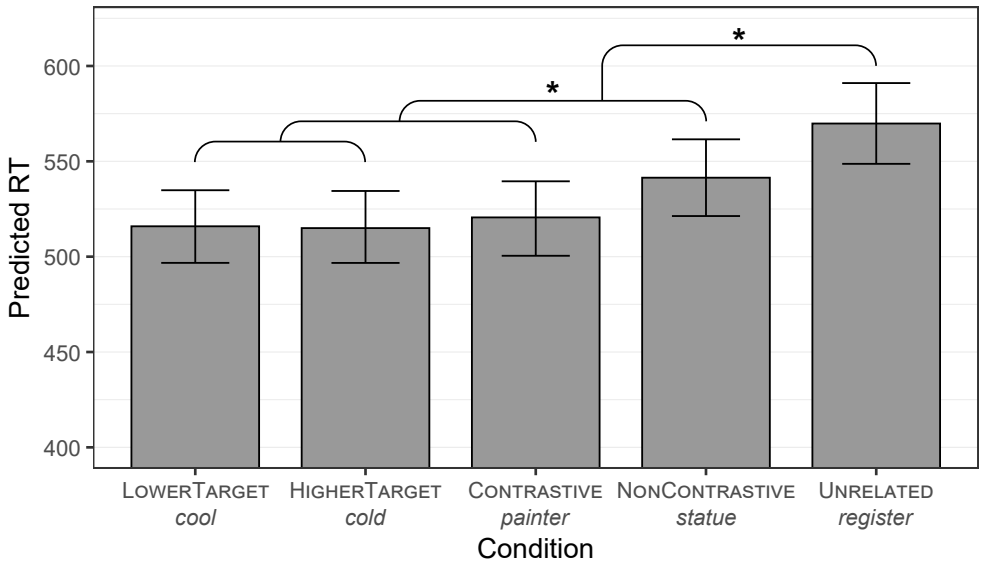

Figure 3: Model predicted RTs with 95% posterior credible intervals. RTs are marginalized over target word length and frequency and experimental block. Credible differences are shown by a $*$.

because RFR is taken to specifically invoke **higher** alternatives, we predict that RFR will not lead to additional facilitation for **lower** alternatives (i.e., *cold* uttered with RFR should not lead to faster lexical decisions of *cool*). In other words, there should be an asymmetry in the processing profile of RFR given the **directional** relationship between the auditory prime and the visual target. We focus on only the critical LOWERTARGET and HIGHERTARGET conditions, again using a Bayesian lognormal distributional model with an added predictor of tune and its interaction with condition.[6]

Figure 4 shows the posterior predicted differences, in terms of percent change ($\%\Delta$), of each tune-condition combination to the condition means previously seen in Figure 3. We can observe that, overall, intonation does not have any notable affect within the LOWERTARGET condition, which is in line with the prediction that RFR does not modulate activation of the lower alternative. Within the HIGHERTARGET condition, the falling tunes are also not credibly different from the HIGHERTARGET condition mean. We do find credible evidence of additional facilitation for H*L-H% ($\hat{\beta} = -0.019, \%\Delta = -1.86\%, CI = [-0.035, -0.003]$), but **less** facilitation for L*HL-H% ($\hat{\beta} = 0.019, \%\Delta = 1.96\%, CI = [0.003, 0.036]$). LH*L-H%, which lies in-between H*L-H% and L*HL-H% in phonetic space, does not show credible evidence for a difference from the conditional mean. Generally, that any of the RFR-shaped tunes (here, 2 of them) show an asymmetry such that a change in processing is observed in the HIGHERTARGET condition but not the LOWERTARGET condition is in line with predictions that RFR is associated with higher alternatives. Yet, unlike in the SI rate data in Exp. 1, here we do not observe a pattern wherein all RFR-shaped tunes behave alike; rather, focusing on H*L-H% and L*HL-H%, we have two tunes within the same

[6]Here, condition is treatment coded with HIGHERTARGET as the reference level (coded 0) and LOWERTARGET as the comparison level (coded 1). Tune is a 6-level predictor that is sum coded (H*L-L% used as the reference level, coded -1, and comparison levels coded as +1). Together, these contrast schemes yield fixed effect estimates that reflect deviations of each tune from the HIGHERTARGET condition's mean.

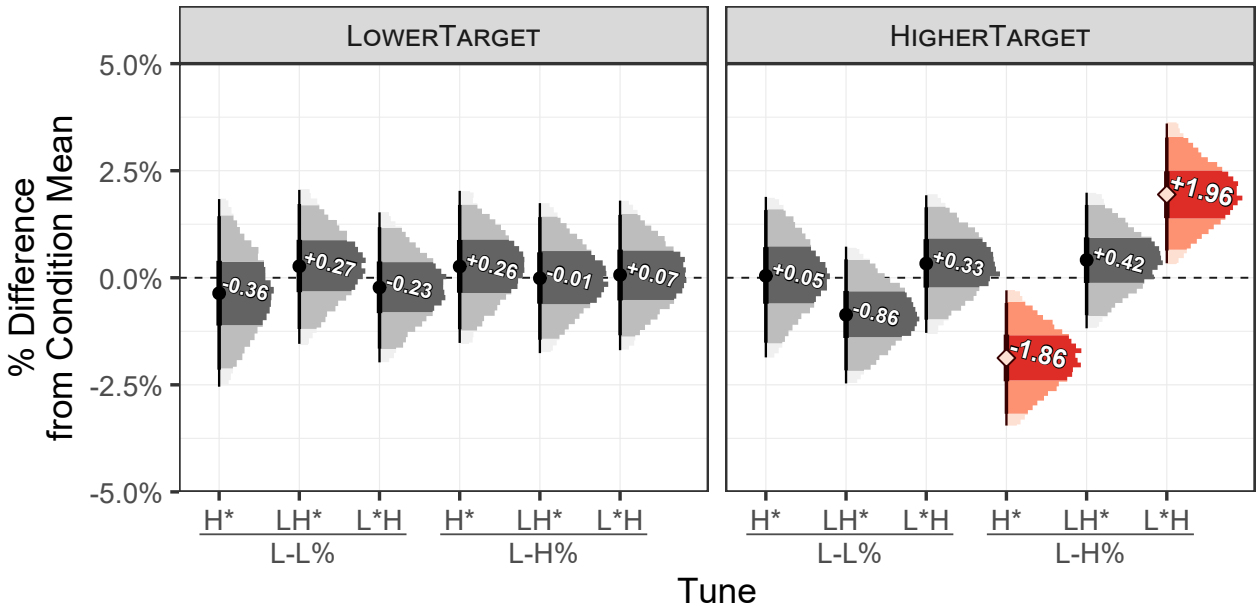

Figure 4: Posterior distribution of RT percent changes with 50%, 89%, and 95% mean highest density intervals. Tunes that are not credibly different from the condition means are shown in gray (means=black circles) while tunes that are credibly different are shown in red (light diamonds).

broad class behaving differently. Regardless, the results of this online task are directly in line with the common theme throughout the RFR literature that RFR is related to higher alternatives, thus providing psycholinguistic evidence for this claim.

5. General Discussion. We investigated whether RFR-shaped tunes behave differently from one another as compared to falling tunes using the same three pitch accents. In our offline inference task, we found a primary distinction between RFR-shaped tunes and falls, with the former encouraging the computation of SI-enriched interpretations. This replicates prior empirical findings on RFR in the context of SI computation (de Marneffe & Tonhauser 2019, Ronai & Göbel to appear) with not only a greater variety of intonational tunes, but also with dialogue contexts that do not overtly mention the higher alternative. (Note that Buccola & Goodhue (2023) find different results, which may be attributable to differences in the experimental paradigms —see Buccola & Goodhue (2023) and Ronai & Göbel (to appear) for further discussion.) With regard to formal pragmatic accounts of RFR, our results are in line with proposals from Göbel (2019) and Göbel & Wagner (2023) where use of RFR enhances the salience of the higher alternatives. These results are incompatible with accounts of RFR which would predict a reduction in the likelihood of SI computation such as Ward & Hirschberg (1985) and Wagner et al. (2013) but may be compatible with other accounts such as those offered by Constant (2012) and Westera (2019), which do not unambiguously predict either an increase or decrease in SI computation.

In our online cross-modal priming with lexical decision task, we find an asymmetry such that two RFR-shaped tunes show different processing signatures when probing a higher scalemate like *cold* compared to a lower scalemate like *cool*. Interestingly, while we hypothesized that increased likelihood of SI would lead to increased facilitation of the higher alternative, we only find this for

one RFR-shaped tune, H*L-H%, which was not the RFR that yielded the highest SI rate. Moreover, the RFR that did yield the highest SI rate instead showed less facilitation of the higher alternative.[7] Our two sets of results seem to paint different pictures: RFR-shaped tunes behave alike in offline interpretation but behave (counterintuitively) differently in online processing.

We can reconcile these results by considering variation at the level of the holistic tune, rather than between-category variation at the level of pitch accents. Our priming results show a difference in the amount of facilitation between the RFR-shaped tunes that use H* (≈2% faster) versus L*+H (≈2% slower), with L+H* lying somewhere in the middle— so we still need to contrast, minimally, H*L-H% with L*+HL-H%. We interpret this pattern a similarly to Hirschberg & Ward (1992), who report that increased scaling in the pitch range of RFR is more likely to yield an incredulous interpretation. More broadly, it is known that increased scaling of pitch range and pitch accent excursions lead to higher conveyance of speaker affect, engagement, and arousal (Ladd et al. 1985, Gussenhoven 2004). While our materials are distinct in their tonal specification (e.g., H*L-H%), they also co-vary in terms of their pitch range; L*+H is known to be more prominent, with higher and later-aligned peak F0 targets, compared to L+H* and H* (Iskarous et al. 2024). Accordingly, we can recast our RFR-shaped tunes under a broad RFR class (based on our SI rates, where the RFR-shaped tunes behaved similarly) with meaningful phonetic variation between a low-scaled RFR (our H*L-H%) and a high-scaled RFR (our L*+HL-H%). When interpreting a high-scaled RFR, additional competing inferences beyond SI (such as incredulity or other particularized inferences) may become more likely. In our priming task, this allows for an interpretation whereby RFR evokes higher alternatives, leading to facilitation, but further increasing the scaling of RFR invites additional competing inferences which strains processing and may mask potential facilitation effects, resulting in less facilitation. In our inference task, the response options are constrained to specifically probe SI computation, allowing for differences in SI rates to arise.

6. Conclusions. We presented results from offline interpretation and online processing experiments, finding evidence for within-category variation at the level of broad Falling and Rise-Fall-Rise tune shapes. In offline interpretation, we found that RFR-shaped tunes increase the likelihood of SI computation compared to falls in indirect question-answer dialogue contexts. In online processing, we found that scalar alternatives behave similarly to contrastive focus alternatives. Moreover, we find an asymmetry when RFR is used with a lower alternative, leading to either additional facilitation when the RFR is low-scaled (H*) or less facilitation when the RFR is high-scaled (L*+H). Neither effect is found when probing the lower alternative after RFR is used with a higher alternative, which rules out the interpretation that the facilitation effect we do find for higher alternatives can be reduced to an effect of semantic priming. Altogether, our results therefore provide evidence of a psycholinguistic correlate to the recurring theme in the literature that RFR evokes higher alternatives. One limitation of our priming experiment is that it is not known whether or not a participant computed the SI-enriched interpretation, precluding directly linking the presence of

[7] Independently, Lacina & Gotzner (2024) report a similar counterintuitive pattern in a text-based priming paradigm, where participants made a lexical decision on the higher alternative following rapid serial visual presentation of a sentence containing a lower alternative. The authors found an inverse correlation between SI rates and facilitation across lexical scales such that scales with higher SI rates displayed less facilitation of their higher alternatives; in other words, the priming results showed the inverse of the typical scalar diversity cline of SI rates.

activation for alternatives to SI calculation. In ongoing work, we address this issue by combining cross-modal lexical decision and inference tasks in a dual-task setup.

References

Aparicio, Helena & Eszter Ronai. 2023. Scalar implicature rates vary within and across adjectival scales. In *Semantics and linguistic theory*, 110–130. 10.3765/t7t8pn98.

Arvaniti, Amalia. 2019. Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. In *Proceedings of the international congress of the phonetic sciences*, Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Balota, David A, Melvin J Yap, Keith A Hutchison, Michael J Cortese, Brett Kessler, Bjorn Loftis, James H Neely, Douglas L Nelson, Greg B Simpson & Rebecca Treiman. 2007. The english lexicon project. *Behavior research methods* 39. 445–459. 10.3758/bf03193014.

Boersma, Paul & David Weenink. 2020. Praat: doing phonetics by computer [computer program]. version 6.2,14.

Braun, Bettina & Lara Tagliapietra. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7-9). 1024–1043.

Buccola, Brian & Daniel Goodhue. 2023. The effect of intonation on scalar and ignorance inferences. In *The proceedings of the chicago linguistic society*, vol. 59, 1–12.

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics and Philosophy* 26(5). 511–545. 10.1023/a:1025887707652.

Bürkner, Paul-Christian. 2021. Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100(5). 1–54. 10.18637/jss.v100.i05. R package version 2.21.6.

Cole, Jennifer, Jeremy Steffman, Stefanie Shattuck-Hufnagel & Sam Tilsen. 2023. Hierarchical distinctions in the production and perception of nuclear tunes in american english. *Laboratory Phonology* 14(1). 10.16995/labphon.9437.

Constant, Noah. 2012. English rise-fall-rise: a study in the semantics and pragmatics of intonation. *Linguistics and Philosophy* 35(5). 407–442. 10.1007/s10988-012-9121-1.

Göbel, Alexander. 2019. Additives pitching in: L*+H signals ordered focus alternatives. In *Semantics and linguistic theory*, vol. 29, 279–299. 10.3765/salt.v29i0.4612.

Göbel, Alexander & Michael Wagner. 2023. On a concessive reading of the rise-fall-rise contour: contextual and semantic factors. *Experiments in Linguistic Meaning* 2. 83–94.

Gotzner, Nicole, S Solt & A Benz. 2018. Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology* 9(SEP). 1–13. 10.3389/fpsyg.2018.01659.

Gotzner, Nicole, Isabell Wartenburger & Katharina Spalek. 2016. The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition* 8(1). 59–95.

Grice, Herbert P. 1975. Logic and conversation. In *Speech acts*, 41–58. Brill.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge University Press.

Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in english. *Journal of phonetics* 20(2). 241–251. 10.1016/S0095-4470(19)30625-4.

Horn, Laurence Robert. 1972. *On the semantic properties of logical operators in english*: University of California, Los Angeles dissertation.

Husband, E Matthew & Fernanda Ferreira. 2016. The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2). 217–235.

Iskarous, Khalil, Jennifer Cole & Jeremy Steffman. 2024. A minimal dynamical model of intonation: Tone contrast, alignment, and scaling of american english pitch accents as emergent properties. *Journal of Phonetics* 104. 101309. 10.1016/j.wocn.2024.101309.

Lacina, Radim & Nicole Gotzner. 2024. Exploring scalar diversity through priming: A lexical decision study with adjectives. In *Proceedings of the annual meeting of the cognitive science society*, vol. 46, https://escholarship.org/uc/item/4mr322mh.

Ladd, D Robert, Kim EA Silverman, Frank Tolkmitt, Günther Bergmann & Klaus R Scherer. 1985. Evidence for the independent function of intonation contour type, voice quality, and f 0 range in signaling speaker affect. *The Journal of the Acoustical Society of America* 78(2). 435–444.

de Marneffe, Marie-Catherine & Judith Tonhauser. 2019. Inferring meaning from indirect answers to polar questions: The contribution of the rise-fall-rise contour. In *Questions in discourse*, 132–163. Brill. 10.1163/9789004378322_006.

Peirce, Jonathan W. 2007. Psychopy—psychophysics software in python. *Journal of neuroscience methods* 162(1-2). 8–13. 10.1016/j.jneumeth.2006.11.017.

Pierrehumbert, J. 1980. *The phonology and phonetics of english intonation*: Massachusetts Institute of Technology dissertation.

Ronai, Eszter & Alexander Göbel. to appear. Watch your tune! on the role of intonation for scalar diversity. *Glossa Psycholinguistics* .

Ronai, Eszter & Ming Xiang. 2021. Exploring the connection between Question Under Discussion and scalar diversity. *Proceedings of the Linguistic Society of America* 6(1). 649.

Ronai, Eszter & Ming Xiang. 2023. Tracking the activation of scalar alternatives with semantic priming. *Experiments in Linguistic Meaning* 2. 229–240. 10.3765/elm.2.5371.

Ronai, Eszter & Ming Xiang. 2024. What could have been said? Alternatives and variability in pragmatic inferences. *Journal of Memory and Language* 136. 104507.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1(1). 75–116.

Steffman, Jeremy, Jennifer Cole & Stefanie Shattuck-Hufnagel. 2024. Intonational categories and continua in american english rising nuclear tunes. *Journal of Phonetics* 104. 101310.

Van Tiel, Bob, Emiel Van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of semantics* 33(1). 137–175. 10.1093/jos/ffu017.

Wagner, Michael, Elise McClay & Lauren Mak. 2013. Incomplete answers and the rise-fall-rise contour. In *Proceedings of the 17th workshop on the semantics and pragmatics of dialogue*, 140–149. http://semdial.org/anthology/Z13-Wagner_semdial_0018.pdf.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61. 747–776. 10.2307/414489.

Watson, Duane G, Michael K Tanenhaus & Christine A Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. l+ h. *Cognitive science* 32(7). 1232–1244.

Westera, Matthijs. 2019. Rise-fall-rise as a marker of secondary QUDs. In *Secondary content*, 376–404. Brill. 10.1163/9789004393127_015.

Yan, Mengzhu & Sasha Calhoun. 2019. Priming effects of focus in mandarin chinese. *Frontiers in Psychology* 10. 10.3389/fpsyg.2019.01985.