Contra assertions, feedback improves word recognition: How feedback and lateral inhibition sharpen signals over noise

James S. Magnuson[1,2,3], Anne Marie Crinnion[1], Sahil Luthra[4], Phoebe Gaston[1], and Samantha Grubb[1]

[1]University of Connecticut. Storrs, CT, USA

[2]BCBL. Basque Center on Cognition Brain and Language, Donostia-San Sebastián, Spain

[3]Ikerbasque. Basque Foundation for Science, Bilbao, Spain

[4]Carnegie Mellon University, Pittsburgh, PA, USA

Author Note

While revising this article, we were saddened to learn of Anne Cutler's passing. We learned so much from her endlessly insightful work and her contributions to the spirited debate that this article continues. This research was supported in part by U.S. National Science Foundation grants BCS-PAC 1754284, BCS-PAC 2043903, and NRT 1747486 (PI: JSM). This research was also supported in part by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S and through project PID2020-119131GB-I00 (BLIS). SL was supported by an NSF Graduate Research Fellowship and by NIH NRSA 1F32DC020625-01. AMC and PG were supported by NIH T32 DC017703 (E. Myers and I.-M. Eigsti, PIs). We thank Arty Samuel, Jay McClelland, Thomas Hannagan, and Rachel Theodore for comments that improved this work. Correspondence concerning this article should be addressed to James Magnuson, Department of Psychological Sciences, University of Connecticut, Storrs, CT, 06069-1020, USA. E-mail: james.magnuson@uconn.edu

Abstract

Whether top-down feedback modulates perception has deep implications for cognitive theories. Debate has been vigorous in the domain of spoken word recognition, where competing computational models and agreement on at least one diagnostic experimental paradigm suggest that the debate may eventually be resolvable. Norris and Cutler (2021) revisit arguments against lexical feedback in spoken word recognition models. They also incorrectly claim that recent computational demonstrations that feedback promotes accuracy and speed under noise (Magnuson, Mirman, Luthra, Strauss, & Harris, 2018) were due to the use of the Luce choice rule rather than adding noise to inputs (noise was in fact added directly to inputs). They also claim that feedback cannot improve word recognition because feedback cannot distinguish signal from noise. We have two goals in this paper. First, we correct the record about the simulations of Magnuson et al. (2018). Second, we explain how interactive activation models selectively sharpen signals via joint effects of feedback and lateral inhibition that boost lexically-coherent sublexical patterns over noise. We also review a growing body of behavioral and neural results consistent with feedback and inconsistent with autonomous (non-feedback) architectures, and conclude that parsimony supports feedback. We close by discussing the potential for synergy between autonomous and interactive approaches.

Contra assertions, feedback improves word recognition:
How feedback and lateral inhibition sharpen signals over noise

Introduction

The decades-old feedback debate in spoken word recognition (SWR) has significant implications for perception and cognition. It echoes similar debates in vision (Firestone & Scholl, 2016) and cognition more generally (Clark, 2013). At stake are longstanding questions about modularity (Fodor, 1983) and cognitive penetrability of perception (Pylyshyn, 1999), with implications for understanding the cognitive and neurobiological bases of perception. The domain of SWR stands out because the debate has revolved around predictions from implemented computational models – primarily the interactive activation model, TRACE (McClelland & Elman, 1986) vs. autonomous models without feedback (primarily Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000) – and a crucial experimental paradigm that both sides agree could provide definitive evidence supporting the need for feedback (*lexically-mediated compensation for coarticulation*; Elman & McClelland, 1988). This suggests that the debate in SWR has high potential to be resolved, which would inform theories of perception and cognition more generally.

This article was motivated by an erroroneous claim in Norris and Cutler (2021), who revisit the feedback debate in SWR and devote substantial text to a critique of Magnuson et al. (2018). Magnuson et al. made a comprehensive case for how feedback can improve SWR that included simulations (with TRACE; McClelland & Elman, 1986) demonstrating that feedback promotes faster and more accurate SWR as noise is added to inputs. Norris and Cutler (2021) claim that Magnuson et al. (2018)'s results were an artifact of a flawed "workaround" in TRACE: emulating noise at a decision stage rather than adding noise directly to inputs. Norris and Cutler were mistaken: Magnuson et al. (2018) clearly describe adding noise to model *inputs*.

However, we do more in this article than simply correct the record on this point. We will review the case against feedback raised by Norris, McQueen, and Cutler (2018), and we shall see that empirical and computational results contradict this case, and that feedback is more parsimonious with extant behavioral and neural data. We will also explain how joint effects of feedback and lateral inhibition allow interactive activation models to selectively enhance signals over noise – the key to understanding how feedback improves word recognition under noise. We conclude with suggestions for moving beyond the feedback debate, and using the tension between autonomous and interactive theories to drive synergistic advances in understanding SWR.

The case against feedback in spoken word recognition

The case against feedback in SWR consists of four primary points.

1. Models without feedback should be preferred because *feedforward systems are simpler* (Norris et al., 2000, pp. 299, 323).
2. *Feedback would hinder processing:* Mixing top-down and bottom-up information makes veridical perception impossible and implies hallucination (Norris et al., 2000, p. 302).
3. *Feedback is not necessary*: Lexical influences on sublexical tasks can be simulated by adding a post-perceptual phoneme decision pathway to a feedforward system (Norris et al., 2000).
4. Even in interactive activation models, *feedback serves no useful purpose* other than fitting data (lexical influences on phoneme tasks; Norris & Cutler, 2021; Norris et al., 2000).

We disagree with each point. We are motivated to respond to Norris and Cutler (2021) because their argument depends crucially upon a mistaken assertion about Magnuson et al. (2018)'s simulations and misunderstandings about the Luce Choice Rule and dynamics of interactive activation models. We first explain why the four points above are invalid, and provide a detailed explanation of how feedback in interactive activation models – in concert with lateral inhibition – promotes accuracy and speed. We conclude with discussions of parsimony, and the relative utility and potential complementarity of ideal observer (e.g., Shortlist B; Norris & McQueen, 2008) and neurally-inspired algorithmic models (e.g., TRACE; McClelland & Elman, 1986).

Is the case against feedback supported?

Are systems without feedback simpler? This assertion could only be true if a feedforward system could account for everything a feedback system can *without additional mechanisms*, or by *adding a simpler mechanism*. However, a purely feedforward system cannot account for lexical effects on sublexical decisions (e.g., Ganong, 1980; Rubin, Turvey, & Van Gelder, 1976; Samuel, 1981, 1996, 2001). Norris et al. (2000) added a special-purpose postperceptual decision mechanism to their *Merge* model to account for such effects without feedback. This is at least as complex as adding feedback (Figure 1): It requires duplicating the sublexical layer and adding two sets of weights; feedback adds one layer of weights and no nodes.

Would feedback hinder perception? Norris and Cutler (2021) assert that feedback necessarily hinders perception. However, interactive activation models are readily parameterized to provide strong

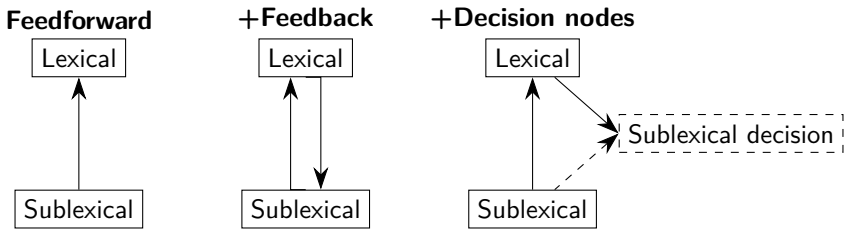

Figure 1. Comparing the complexity of purely feedforward architecture (left) with the addition of feedback (center), as in TRACE, or sublexical decision nodes (right), as in Merge (Norris et al., 2000). Either feedback connections or decision nodes are required to account for lexical effects on sublexical decisions. The decision-node architecture requires more additional nodes and connections than feedback (dashed lines). *Reproduced from Magnuson (2022b).*

bottom-up priority – as Magnuson et al. (2018) discuss (p. 12; see also Magnuson, Mirman, & Harris, 2012; Magnuson, Mirman, & Myers, 2013; McClelland & Elman, 1986) – while also accurately simulating human decision delays and misperceptions (McClelland, Mirman, Bolger, & Khaitan, 2014; Mirman, McClelland, & Holt, 2005). The original TRACE parameters achieve a balance sufficient to simulate dozens of aspects of human speech perception and SWR while enforcing strong bottom-up priority (Magnuson & Crinnion, 2022). Note that this is a remarkable aspect of the model; it is unusual for a complex model to generalize to many phenomena with fixed parameters.

Is feedback necessary in theories of spoken word recognition? While a postperceptual sublexical decision mechanism (Figure 1, right panel) can simulate lexical influences on some sublexical tasks, there are crucial exceptions. Samuel (1997) makes a compelling case that lexically-driven selective adaptation from phonemes replaced with noise supports interaction, but Norris et al. (2000) argue the results could actually manifest at a lexical rather than sublexical level. While we disagree, it is the case that both sides agree that *lexically-mediated compensation for coarticulation (LCfC)* (Elman & McClelland, 1988) provides a diagnostic test for distinguishing feedforward and feedback accounts (Norris, McQueen, & Cutler, 2016; Pitt & McQueen, 1998).

In LCfC, the question is whether a lexically-restored phoneme can drive a sublexical coarticulatory effect (compensation for coarticulation; Mann & Repp, 1981), which would be consistent only with feedback (on the logic that CfC involves a prelexical, phonetic-level interaction, and so lexical influence on CfC would constitute what McClelland et al., 2014, call a 'knock-on consequence' of interaction). Previous positive results (e.g., Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003) have been called into question due to replication failures (McQueen, Jesse, & Norris, 2009), and evidence that transitional probabilities in nonwords can also drive phoneme restoration that can drive CfC in the absence of lexical context (Pitt & McQueen, 1998). Recently,

though, Luthra et al. (2021) observed that few studies verified that items could separately drive phoneme restoration and CfC before being combined in LCfC; if items cannot drive the component effects separately, they will not drive LCfC when combined. Luthra et al. (2021) reported robust, replicable LCfC effects when they only used materials capable of generating phoneme restoration and compensation for coarticulation *separately* before being combined in LCfC (including materials where transitional probability and lexical context had opposite biases).[1]

There is remarkably strong neurophysiological evidence from monkeys that feedback connections from higher levels in the visual system 'serve to amplify and focus' responses in lower levels (Hupé et al., 1998). There are numerous neural studies that suggest feedback supports online human language processing. Setting aside cases that Norris et al. (2016) argue are not sufficiently specific (though we do not agree with this assessment; e.g., Gow & Olson, 2015; Gow, Segawa, Ahlfors, & Lin, 2008; Myers & Blumstein, 2008), recent ERP evidence for early impact of lexical context suggests feedback is necessary (Getz & Toscano, 2019; Noe & Fischer-Baum, 2020).

In light of this new evidence, we conclude that (a) **feedback is necessary** and (b) **autonomous alternatives** to feedback (Norris & McQueen, 2008; Norris et al., 2000) **are insufficient**: they cannot account for lexically-mediated compensation for coarticulation or early lexical modulation of low-level neural responses because they prohibit lexical-level information from interacting with sublexical computations.

Does feedback have a useful purpose? The assertion that *feedback serves no useful purpose* follows from a misunderstanding about interactive activation and depends on two erroneous claims and one outdated claim.

i. ***Claim****: Feedback cannot distinguish signal from noise*, so feedback can only 'reinforce the status quo' (Norris & Cutler, 2021, pp. 3-4). Therefore, the claim is that the system can do no better than to select the word with best fit to the bottom-up input (Norris et al., 2000, p. 301), and it would be impossible for feedback to improve upon this, because feedback can only mirror input (whether signal or noise).

ii. ***Claim****: Feedback in TRACE does not promote faster word recognition.* In TRACE simulations comparing recognition times for 21 specially-selected words with feedback on or off (Frauenfelder & Peters, 1998), as many words were recognized more quickly *without* feedback as were recognized more quickly *with* feedback (cited as critical support by Norris et al., 2000, p. 302, p. 324).

[1] McQueen, Jesse, and Mitterer (2023) argue that some unspecified combination of stimulus flaws, transitional probabilities, and unknown factors could drive these new LCfC data. As Luthra, Crinnion, Saltzman, and Magnuson (submitted) reply, though, this case is ad hoc, whereas interaction provides a coherent and parsimonious explanation.

iii. ***Claim**: Feedback does not promote faster or more accurate word recognition in noise.* Norris and Cutler (2021) (mistakenly) claim simulations in Magnuson et al. (2018) only *appear* to show feedback advantages but are due to a "workaround" of posthoc application of the Luce Choice Rule to activations, rather than direct addition of noise to model inputs.

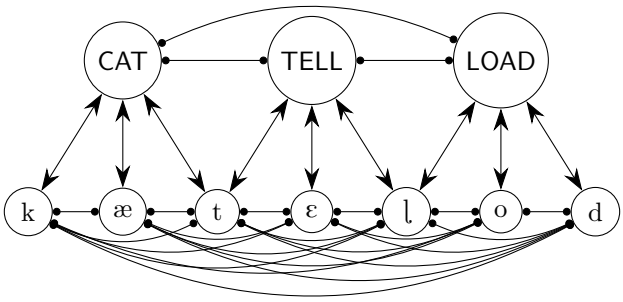

Figure 2. Interactive activation example. Arrows denote excitatory connections (7 input phonemes feed forward to 3 words, which send feedback to constituent phonemes). Edges with bulb connectors indicate lateral inhibition links within layers. *Reproduced from Magnuson (2022a).*

First, consider claim (i), *feedback cannot distinguish signal from noise*. The assumption that feedback cannot selectively reinforce signal over noise neglects *lateral inhibition*. Consider Figure 2. If the input is /tɛl/, TELL would be most activated by the bottom-up signal, CAT and LOAD would be partially activated by the signal, and all three would receive additional activation if noise were added.

Without lateral inhibition, feedback *must* simply mirror the bottom-up combination of signal plus noise. However, so long as TELL has greater bottom-up support than other words, when it becomes even slightly more activated than other words, it begins to inhibit CAT and LOAD, reducing the feedback those nodes send to their constituent phonemes. The subsequently enhanced feedback from TELL promotes greater activation of /t/, /ɛ/, and /l/, shifting the balance of signal and noise already, and lateral inhibition between phonemes further increases that advantage. As excitatory activation iteratively resonates vertically (words ↔ phonemes) and inhibition flows laterally within each layer, the signal will be *sharpened* (Figure 3). When inputs are noisy, signal is promoted and activation due to noise is inhibited due to *iterative refinement*. Typically, this will drive faster target activation and, under noise, better accuracy (as demonstrated via simulations by Magnuson et al., 2018, who observed advantages in proportion correct of ∼0.05 to ∼0.2 at various levels of noise, as well as robust 5-10% advantages in response times for correctly recognized items; see their Figure 3, as well as Figures A1 and A2 below).

Next, consider claim (ii): *feedback does not promote faster word recognition*. In simulations with only

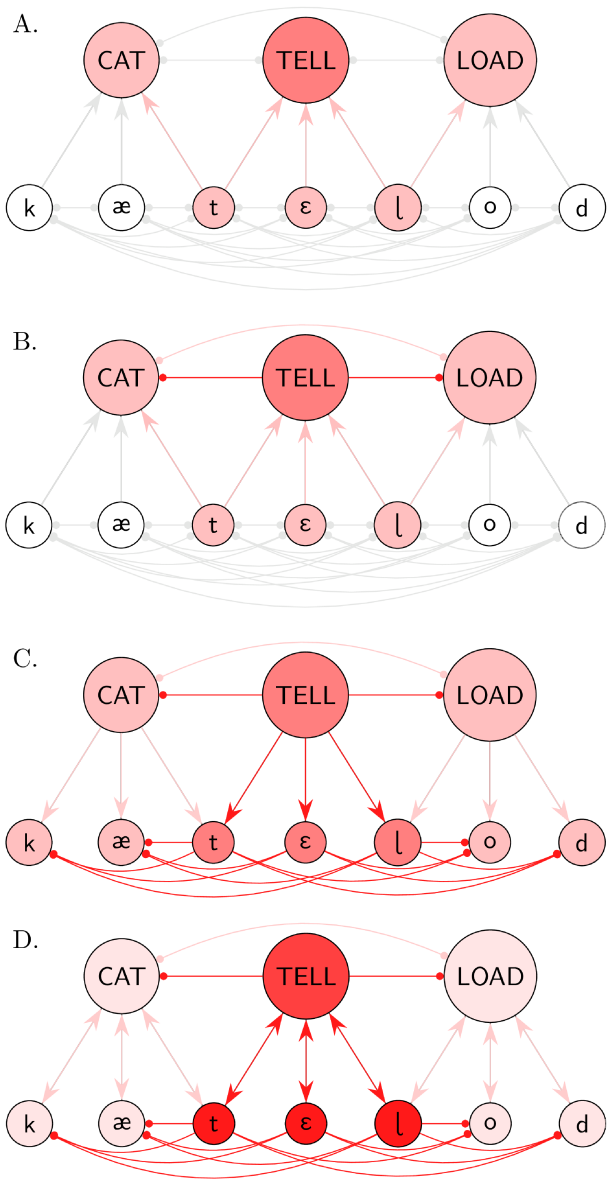

Figure 3. A schematic of iterative refinement (or signal sharpening, cf. Blank & Davis, 2016) in interactive activation. (A) Given input consistent with 'tell', TELL becomes strongly activated while CAT and LOAD become partially activated. (B) TELL inhibits CAT and LOAD. (C) Words send feedback to constituent phonemes, with lateral inhibition at the phoneme level enhancing the advantage for phonemes that are relatively strongly activated. (D) Over subsequent cycles of excitatory and inhibitory activation flow, the activations for TELL and its phonemes are iteratively enhanced / sharpened. Note that if random noise were added to the inputs, the same refinement / sharpening would happen so long as the target word has at least slightly higher activation than other words. The most important connections are highlighted in red in each panel, with darker red indicating greater activation. *Reproduced from Magnuson (2023).*

21 words, as many words were recognized more quickly without feedback as were recognized more quickly with feedback (Frauenfelder & Peters, 1998). However, that result is superseded by Magnuson et al. (2018)'s simulations with hundreds of words. Magnuson et al. found that, even without noise, 57% of words from the original ∼200-word TRACE lexicon were recognized more quickly with feedback and 27% were slower (the rest did not change). With a ∼900-word lexicon, without noise, 38% were recognized more quickly with feedback, while 36% were recognized more slowly. However, for both lexicons, progressively stronger feedback advantages emerged (in both speed and accuracy) as noise increased (with substantial noise, ∼60-80% were recognized more quickly with feedback, and accuracy was substantially higher with feedback than without at even moderate levels of noise).

Finally, claim (iii) is that *feedback advantages are artifactual* (Norris & Cutler, 2021, p. 3):

> ...Can recognition of a degraded signal be improved by the kind of activation feedback incorporated in TRACE? The answer is again no ... The degraded-speech case still regularly causes confusion (e.g., [Magnuson et al., 2018]), and may have its origin in the original version of TRACE, which ruled out noise both in the input and during processing ... (The effect of noise was simulated by adding a constant amount of noise to a decision process – the Luce choice rule – operating on the output of the network.) In such a system, where feedback and noise are operating at different levels, feedback (as part of processing) can alter the relative activation of a word and its competitors but at the same time have no effect at all on any noise (which operates separately on final outputs) ... because of a workaround in the model, simulations using TRACE can give the impression that feedback can improve performance. ... In a real system with degraded input, signal and noise are in the same processed channel, and feeding activation back to the phoneme level will boost both the signal and the noise equally. Feedback can here do nothing to improve the signal to noise ratio.

Norris and Cutler (2021) mistakenly claim Magnuson et al. (2018) simulated noise by "adding a constant amount of noise to a decision process - the Luce choice rule - operating on the output of the network." This is incorrect. Magnuson et al. *did* apply the choice rule to activations to generate predicted response probabilities (standard in many previous simulations; e.g., Frauenfelder & Peters, 1998; McClelland & Elman, 1986), but this had nothing to do with noise. The noise procedure was specified in a subsection labeled **'Noise'** (Magnuson et al., 2018, pp. 3-4); noise was added *before* response probabilities were calculated:

> Gaussian noise was sampled from a normal distribution function and added independently

to each element of the input stimulus vector for each time step (cf. McClelland, 1991).

Since noise was added to inputs, noise propagated forward and backward through the system, and thus *signal and noise **were** in the "same processed channel".* The Luce Choice Rule rescales values (amplifying larger values and squashing smaller values). Thus, it could *amplify* differences present in raw activations, but it could not *generate* differences that were not already there. To assuage potential concerns, we present a replication of Magnuson et al.'s feedback advantages using raw activations (i.e., with no use of the Luce Choice Rule) in the Appendix.[2]

The case against feedback is not supported. In summary, *autonomous models are not simpler* (see Magnuson et al., 2018, for a technical case that the feedforward alternative is more complex); there is *no evidence that feedback necessarily hinders perception* (and there are many demonstration proofs that feedback and feedforward gains can be balanced to provide benefits of feedback while maintaining bottom-up priority); ***feedback is necessary*** to account for (a) robust effects (Luthra et al., 2021) that autonomy proponents agree provide a gold-standard test for feedback (McQueen et al., 2009; Norris et al., 2016; Pitt & McQueen, 1998) and (b) recent evidence for early lexical modulation of low-level ERP responses to speech (e.g., Getz & Toscano, 2019; Noe & Fischer-Baum, 2020). Crucially Magnuson et al. (2018) have demonstrated (and we have replicated in the Appendix) *feedback's useful purpose* (in concert with lateral inhibition): making the system robust against noise. Other potential feedback benefits include predicting/anticipating inputs (Bar, 2003), attentional control, and stabilization (Bonte, Parviainen, Hytönen, & Salmelin, 2005).

Parsimony

Given two explanations, so long as both account for relevant data, Occam counsels us to prefer the simpler one. Proponents of autonomous models assert that SWR models without feedback are "incontrovertibly simpler" (Norris et al., 2016), but have not provided a formal basis for this assertion (Magnuson et al., 2018). Indeed, autonomous models add a special-purpose "decision" pathway outside normal perception to account for lexical effects on sublexical tasks that is *at least* as complex as a feedback pathway (Norris & McQueen, 2008; Norris et al., 2000), and arguably more complex (Magnuson et al., 2018).

[2] McClelland (1991) and Movellan and McClelland (2001) acknowledge that Massaro (1989) was correct that TRACE activations converted to response probabilities via the Luce choice rule violate *logistic additivity*, but they also demonstrate that adding noise to TRACE inputs and/or activations *instead* is preferable because the resulting activations do not violate principles of Bayesian inference; thus, in many cases, modelers should add noise to TRACE (or similar interactive activation models) rather than applying the Luce choice rule. However, to be clear, Magnuson et al. (2018) only applied the Luce choice rule *after* applying noise to inputs.

Autonomous models are increasingly incompatible with relevant data. As discussed, autonomous models cannot explain LCfC (recently demonstrated to be robust when items are rigorously pretested Luthra et al., 2021) or recent ERP evidence for rapid lexical impact on low-level responses to speech (Getz & Toscano, 2019; Noe & Fischer-Baum, 2020). Other examples in speech and SWR appear to require feedback (e.g.: Blank & Davis, 2016; Bonte et al., 2005; Cope et al., 2017; Samuel, 1997; Sohoglu & Davis, 2020; Sohoglu, Peelle, Carlyon, & Davis, 2012). Furthermore, there is pervasive evidence for feedback in vision (e.g.: Bar et al., 2006; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Hupé et al., 1998; McMains & Kastner, 2011; Mechelli, 2004; Zanto, Rubens, Bollinger, & Gazzaley, 2010; Zanto, Rubens, Thangavel, & Gazzaley, 2011; Zhang et al., 2014) and audition (e.g.: Alain, Arnott, & Picton, 2001; Bendixen, SanMiguel, & Schröger, 2012; Davis & Johnsrude, 2007; Elhilali, Xiang, Shamma, & Simon, 2009; Kazimierczak et al., 2022; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002).

On balance, the principle of parsimony favors feedback. For extensive theoretical discussions that extend beyond the domain of SWR, see Bar (2003), Clark (2013), Gilbert and Li (2013), Lupyan (2015), and Spivey (2023).

Discussion

The case that autonomous models should be preferred to feedback does not hold. Behavioral (e.g., Luthra et al., 2021) and neural (e.g., Getz & Toscano, 2019; Noe & Fischer-Baum, 2020) empirical results support sublexical impact of lexical knowledge in ways models without feedback (Norris & McQueen, 2008; Norris et al., 2000) cannot accommodate, and simulations (Magnuson et al., 2018) demonstrate how feedback in concert with lateral inhibition in interactive activation sharpens signals relative to noise (see the Appendix for a replication using raw TRACE activations).

Proponents of interactive and autonomous theories appear, in recent publications, to have different goals. Norris and Cutler (2021) claim that SWR systems cannot make better choices than to select the word with best initial bottom-up match to input. To take this literally would be inconsistent with Bayes' theorem, which is why Shortlist B takes into account both prior probability and word likelihoods when evaluating bottom-up input.

If one is primarily concerned with 'computational adequacy' (achieving the best possible performance; McClelland & Elman, 1986), one could stop at choosing the best item with respect to priors and likelihoods. However, we are interested in 'psychological adequacy' (McClelland & Elman): *how* humans recognize spoken words, with the aim of developing theories that span behavior, cognition, development, and neurobiology (cf. Magnuson et al., 2020). On this view, the criteria for model

comparisons goes beyond how to most simply achieve highest performance, with the aim of providing an explanation most compatible with human behavior and neurobiology; with the goal of advancing theories towards comprehensive accounts of the cognitive and neural bases for human performance. This view also endorses a search for algorithms that approach optimality via *satisficing* (Simon, 1956) algorithms that reduce computational complexity (Gigerenzer & Goldstein, 1991), and may better reflect mechanisms underlying human SWR, which must operate over an ambiguous signal via limited cognitive resources under severe time pressure (cf. McClelland et al., 2014, who explicitly frame their *interactive activation hypothesis* as the basis for humans *to approximate optimality* under the constraints of real-time processing and partial information).

Through this lens, feedback provides two interesting steps beyond an initial rational analysis / ideal observer model. First, while Shortlist B is endowed with lexical knowledge (priors and likelihoods) and the luxury of conditioning interpretation of bottom-up inputs on lexical knowledge (without considering algorithmic or neural constraints), feedback in interactive activation instantiates an implicit generative model that embodies approximate word likelihoods given input patterns (since words feed back to their constituent phonemes), without the need to store a lookup table of likelihoods.[3] While the TRACE architecture is not fully neurally plausible, its "neurally inspired" nature reduces the distance towards developing more realistic models.

Second, feedback in concert with lateral inhibition provides 'signal sharpening' (cf. Blank & Davis, 2016) via a process of *iterative refinement*, schematized in Figure 3. Small bottom-up advantages are amplified by successive cycles of lateral resonance (inhibition) and vertical resonance (feedforward-feedback). As inputs become noisier, response times in our simulations become longer, as it takes more iterations for signals to sharpen sufficiently to meet the decision threshold (and as inputs become noisier, there is a greater chance of misses [targets failing to reach threshold] or false alarms [non-targets exceeding threshold]).

If iterative refinement ideally settles on the word with the best bottom-up fit, how does it yield better results than a system that would just choose the word with best bottom-up fit immediately (cf. Norris et al., 2000)? This follows from having a *decision policy* (see Appendix). Norris (2006) points out that when inputs are ambiguous, trade-offs of time constraints and error costs may modify ideal decision policies. Decision policies (which may involve an absolute threshold, or the relative activation of the most active item to the next-most active item; cf. Hannagan, Magnuson, & Grainger, 2013) can implicitly represent confidence: e.g., more confidence is warranted when the top lexical candidate is marginally more likely than many words with low likelihoods vs. one where the target has the same

[3] Of course, one could respond that, e.g., frequency is 'given' in TRACE as a lookup table. However, phoneme → *word weights proportional to frequency* are consistent with general learning principles (Dahan, Magnuson, & Tanenhaus, 2001).

likelihood but is marginally more likely than a single item (while other items have much lower likelihoods). Indeed, in the former case, iterative refinement via interaction will boost the target more strongly.

With the goal of psychological adequacy, it is productive to consider how models like Shortlist B and TRACE complement one another. Shortlist provides a baseline as an approximate ideal observer model. Shortlist does not (yet) address how such a model could be implemented in a biologically-plausible way, nor how such a system could develop. Shifting from a debate mode to the aim of seeking synergy between autonomous and interactive approaches, we might focus on understanding how human performance and/or neurobiology differ from the ideal observer baseline (which would require developing falsifiable hypotheses from Shortlist), whether models like TRACE differ from the ideal observer baseline in human-like ways (which would be facilitated by developing new falsifiable hypotheses from TRACE), or developing information processing theories at all three of Marr's levels, and comparing theories that start either from assumptions of autonomy or interaction. This might lead us to discover ways to improve or merge the approaches, which would be a welcome shift from the current focus on debate.

While it is important to correct errors and misunderstandings (our motivation for writing this response), the feedback debate must be resolved through empirical studies, model comparisons, and working towards theories that integrate Marr's three levels. This will serve what we assume is a shared goal with our respected opponents in this debate: explaining the cognitive, perceptual, and neurobiological bases of human language processing.

References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1072–1089. Retrieved from http://dx.doi.org/10.1037/0096-1523.27.5.1072 doi: 10.1037/0096-1523.27.5.1072

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. doi: 10.1162/089892903321662976

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. doi: 10.1073/pnas.0507062103

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. Retrieved from http://dx.doi.org/10.1016/j.ijpsycho.2011.08.003 doi: 10.1016/j.ijpsycho.2011.08.003

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fmri patterns during speech perception. *PLoS Biology*, *14(11)*, 1–33.

Bonte, M., Parviainen, T., Hytönen, K., & Salmelin, R. (2005). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex*, *16*(1), 115–123. doi: 10.1093/cercor/bhi091

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi: 10.1017/S0140525X12000477

Cope, T. E., Sohoglu, E., Sedley, W., Patterson, K., Jones, P. S., Wiggins, J., ... Rowe, J. B. (2017). Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nature Communications*, *8*(1). doi: 10.1038/s41467-017-01958-7

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001, Jun). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. doi: 10.1006/cogp.2001.0750

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1-2), 132–147. Retrieved from http://dx.doi.org/10.1016/j.heares.2007.01.014 doi: 10.1016/j.heares.2007.01.014

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. doi: 10.1016/j.cogbrainres.2003.11.010

Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory

scene. *PLoS Biology*, *7*(6), e1000129. doi: 10.1371/journal.pbio.1000129

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27(2)*, 143–165.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, e229. doi: 10.1017/S0140525X15000965

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Frauenfelder, U. H., & Peters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE. In *Localist connectionist approaches to human cognition.* (pp. 101–146). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6(1)*, 110–125. doi: https://doi.org/10.1037/0096-1523.6.1.110

Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science*, *30*(6), 830–841.

Gigerenzer, G., & Goldstein, D. G. (1991). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650-669.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363.

Gow, D., & Olson, B. (2015). Lexical mediation of phonotactic frequency effects on spoken word recognition: A granger causality analysis of mri-constrained meg/eeg data. *Journal of Memory and Language*, *82*, 41–55.

Gow, D., Segawa, J., Ahlfors, S., & Lin, F.-H. (2008). Lexical influences on speech perception: A granger causality analysis of meg and eeg source estimates. *NeuroImage*, *43*, 614–623.

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a trace. *Frontiers in Psychology*, *4:563*. doi: https://doi.org/10.3389/fpsyg.2013.00563

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, *394*, 784-787.

Kazimierczak, K., Craven, A. R., Ersland, L., Specht, K., Dumitru, M. L., Sandøy, L. B., & Hugdahl, K. (2022). Combined fMRI region- and network-analysis reveal new insights of top-down modulation of bottom-up processes in auditory laterality. *Frontiers in Behavioral Neuroscience*, *15*. Retrieved from `http://dx.doi.org/10.3389/fnbeh.2021.802319` doi: 10.3389/fnbeh.2021.802319

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems

are penetrable systems. *Review of Philosophy and Psychology*, *6*, 547-569. doi: 10.1007/s13164-015-0253-4

Luthra, S., Crinnion, A. M., Saltzman, D. S., & Magnuson, J. S. (submitted). Do they know it's Christmash? Lexical knowledge directly impacts speech perception.

Luthra, S., Peraza-Santiago, G., Beeson, K., Saltzman, D., Crinnion, A. M., & Magnuson, J. S. (2021). Robust lexically-mediated compensation for coarticulation: Christmash time is here again. *Cognitive Science*, *45*(4), e12962.

Magnuson, J. S. (2022a). Interactive activation example: Cat, tell, load. Retrieved from `https://doi.org/10.6084/m9.figshare.19306931.v1`

Magnuson, J. S. (2022b). Network complexity schematic: Feedforward vs. feedback vs. merge. Retrieved from `https://doi.org/10.6084/m9.figshare.19306880.v1`

Magnuson, J. S. (2023). Iterative refinement in interactive activation (figures). *figshare*. Retrieved from `https://figshare.com/articles/figure/_/23538378`

Magnuson, J. S., & Crinnion, A. M. (2022). Spoken word recognition. In A. Papafragou, J. Trueswell, & L. Gleitman (Eds.), *The oxford handbook of the mental lexicon* (p. 461-499). doi: 10.1093/oxfordhb/9780198845003.013.23

Magnuson, J. S., & Curtice, A. (in preparation). jsTRACE: A javascript reimplmentation of the TRACE model of speech perception and spoken word recognition.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of christmash past. *Cognitive Science*, *27*, 285–298.

Magnuson, J. S., Mirman, D., & Harris, H. (2012). Computational models of spoken word recognition. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The cambridge handbook of psycholinguistics* (p. 76-103). New York, NY, US: Cambridge University Press.

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology*, *9*(369). doi: 10.3389/fpsyg.2018.00369

Magnuson, J. S., Mirman, D., & Myers, E. (2013). Spoken word recognition. In D. Reisberg (Ed.), *The oxford handbook of cognitive psychology* (p. 412-441). New York, NY, US: Oxford University Press.

Magnuson, J. S., You, H., Luthra, S., Li, M., Nam, H., Escabí, M., ... Rueckl, J. G. (2020). EARSHOT: A minimal neural network model of incremental human speech recognition. *Cognitive Science*, *44*(4), e12823. doi: https://doi.org/10.1111/cogs.12823

Mann, V., & Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 548–558.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech

perception. *Cognitive Psychology*, *21*(3), 398–421. doi: 10.1016/0010-0285(89)90014-5

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, *23*, 1-44. doi: 10.1016/0010-0285(91)90002-6

McClelland, J. L., & Elman, L., J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86. doi: 10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *6*, 1139-1189. doi: 10.1111/cogs.12146

McMains, S., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *The Journal of Neuroscience*, *31*(2), 587–597. doi: 10.1523/jneurosci.3766-10.2011

McQueen, J. M., Jesse, A., & Mitterer, H. (2023). Lexically mediated compensation for coarticulation still as elusive as a white christmash. *Cognitive Science*, *47*(9), e13342. doi: https://doi.org/10.1111/cogs.13342

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical-prelexical feedback during speech perception or: Is it time to stop playing those christmas tapes? *Journal of Memory and Language*, *61*(1), 1–18.

Mechelli, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, *14*(11), 1256–1265. doi: 10.1093/cercor/bhh087

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, *52*, 424-443. doi: 10.1016/j.jml.2005.01.006

Movellan, J. R., & McClelland, J. L. (2001). The morton-massaro law of information integration: Implications for models of perception. *Psychological Review*, *108*(1), 113–148. doi: 10.1037/0033-295X.108.1.113

Myers, E., & Blumstein, S. (2008). The neural bases of the lexical effect: An fmri investigation. *Cerebral Cortex*, *18*(2), 278–288.

Noe, C., & Fischer-Baum, S. (2020). Early lexical influences on sublexical processing in speech perception: Evidence from electrophysiology. *Cognition*, *197*, 1–14.

Norris, D., & Cutler, A. (2021). More why, less how: What we need from models of cognition. *Cognition*, *213*. doi: https://doi.org/10.1016/j.cognition.2021.104688

Norris, D., & McQueen, J. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115(2)*, 357-395. doi: 10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, *23(3)*, 299-370. doi: 10.1017/s0140525x00003241

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

Norris, D., McQueen, J. M., & Cutler, A. (2018). Commentary on "interaction in spoken word recognition models". *Frontiers in Psychology*, *9*, 1-3. doi: 10.3389/fpsyg.2018.01568

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation meditated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365. doi: 10.1017/S0140525X99002022

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, *19*, 394-398.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474-494.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*, 28-51.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*, 348-351.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*(2), 416–434.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–38.

Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *ELife*, *9*, e58077.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. doi: 10.1523/jneurosci.5069-11.2012

Spivey, M. J. (2023). Cognitive science progresses toward interactive frameworks. *Topics in Cognitive Science*, *15*(2), 219–254. doi: 10.1111/tops.12645

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22–29. Retrieved from `http://dx.doi.org/10.1016/j.heares.2009.12.021` doi: 10.1016/j.heares.2009.12.021

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can

modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, *13*(3), 393–405. Retrieved from http://dx.doi.org/10.1016/S0926-6410(01)00131-8 doi: 10.1016/s0926-6410(01)00131-8

Zanto, T. P., Rubens, M. T., Bollinger, J., & Gazzaley, A. (2010). Top-down modulation of visual feature processing: The role of the inferior frontal junction. *NeuroImage*, *53*(2), 736–745. doi: 10.1016/j.neuroimage.2010.06.012

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656–661. doi: 10.1038/nn.2773

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., ... Dan, Y. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science*, *345*(6197), 660–665. doi: 10.1126/science.1254126

Appendix

Simulation: Feedback and activations under noise

Procedure. In this simulation, we use the same approach as Magnuson et al. (2018) to compare word recognition accuracy and recognition time in TRACE – with the important alteration that we use raw activations rather than response probabilities.

We conducted simulations using jsTRACE, a recent reimplementation of TRACE in JavaScript (Magnuson & Curtice, in preparation). jsTRACE performs identically to jTRACE but updates near-obsolete aspects of the Java implementation and provides the ability to do batch scripting via external JavaScript code. We used the default *slex* TRACE lexicon, consisting of 212 words (as well as the "silence" word used to represent a state of no input; the silence word was not included in analyses). We used three levels of feedback (0.00, 0.015, and 0.03, the latter being the default level with small lexicons in TRACE). We combined each level of feedback with seven levels of Gaussian noise (with mean of zero and standard deviation ranging from 0.0 to 1.5 in steps of 0.25). A value sampled from the distribution was added independently to each cell of the input matrix prior to the simulation. To ensure that results under noise were robust, we conducted 10 simulations of every word in the lexicon at all levels of noise greater than zero. We allowed simulations to run for 100 time steps (cycles) in TRACE.

Decision policy. To assess how quickly a model "identifies" a word, we need a decision policy. We cannot simply take the maximum value, as a target's activation may continue increasing throughout an entire simulation. We use a simple threshold-based policy, where a correct identification is defined as the target reaching or exceeding that threshold and no other item reaching it. Recognition time is the cycle where the target's activation first reaches or exceeds the threshold. We first identified the activation threshold that would maximize accuracy for zero feedback without noise; this was 0.4. We then applied that threshold to every simulation (that is, at every level of feedback and noise). Note that any potential bias in this policy favors simulations without feedback, since the threshold optimizes accuracy with zero noise and zero feedback.

Crucially, all analyses were applied to raw activations. We did not transform activations to response probabilities.

Results

In Figure A1, we see a clear replication of the results of Magnuson et al. (2018) based on raw activations. Recognition time is faster with feedback than without at every level of noise – with the exception of $sd = 1.50$, where only 3 correct trials (out of 2120 trials) occurred with feedback set to 0.0

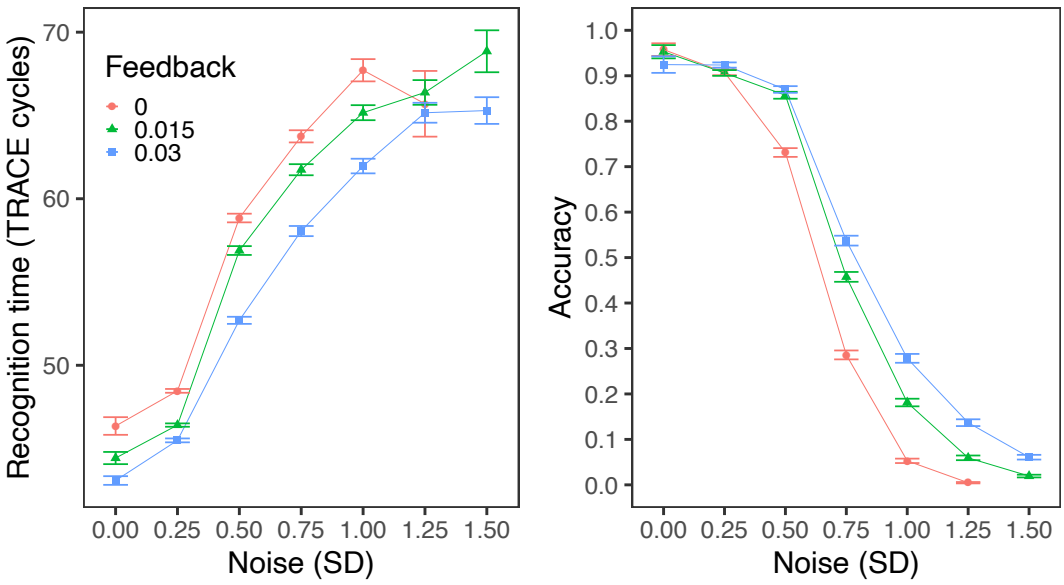

Figure A1. Results of Simulation 1. Replication of Magnuson et al. (2018) using activations instead of response probabilities. Each point represents the outcome of simulating every word in the 212-word *slex* lexicon, with 10 simulations conducted with each word at each noise level greater than zero. The recognition threshold was set to 0.4, which maximized accuracy with feedback and noise set to zero.

(compared to 37 correct trials with feedback at 0.015 and 169 correct trials with feedback at 0.03), rendering this point uninterpretable. Feedback allows higher accuracy once noise is added.

In Figure A2, we plot recognition times for words that were correctly recognized with feedback at the default value of 0.03 and without feedback (set to 0.0) at each noise level. Again, we repeated the simulation of each word 10 times at each level of noise greater than zero. A clear, consistent advantage is observed for the majority of words at each level of noise.

Conclusions

Contra assertions by Norris and Cutler (2021), feedback promotes more robust word recognition performance, and the results reported by Magnuson et al. (2018) were not an artifact of using response probabilities rather than raw activations.

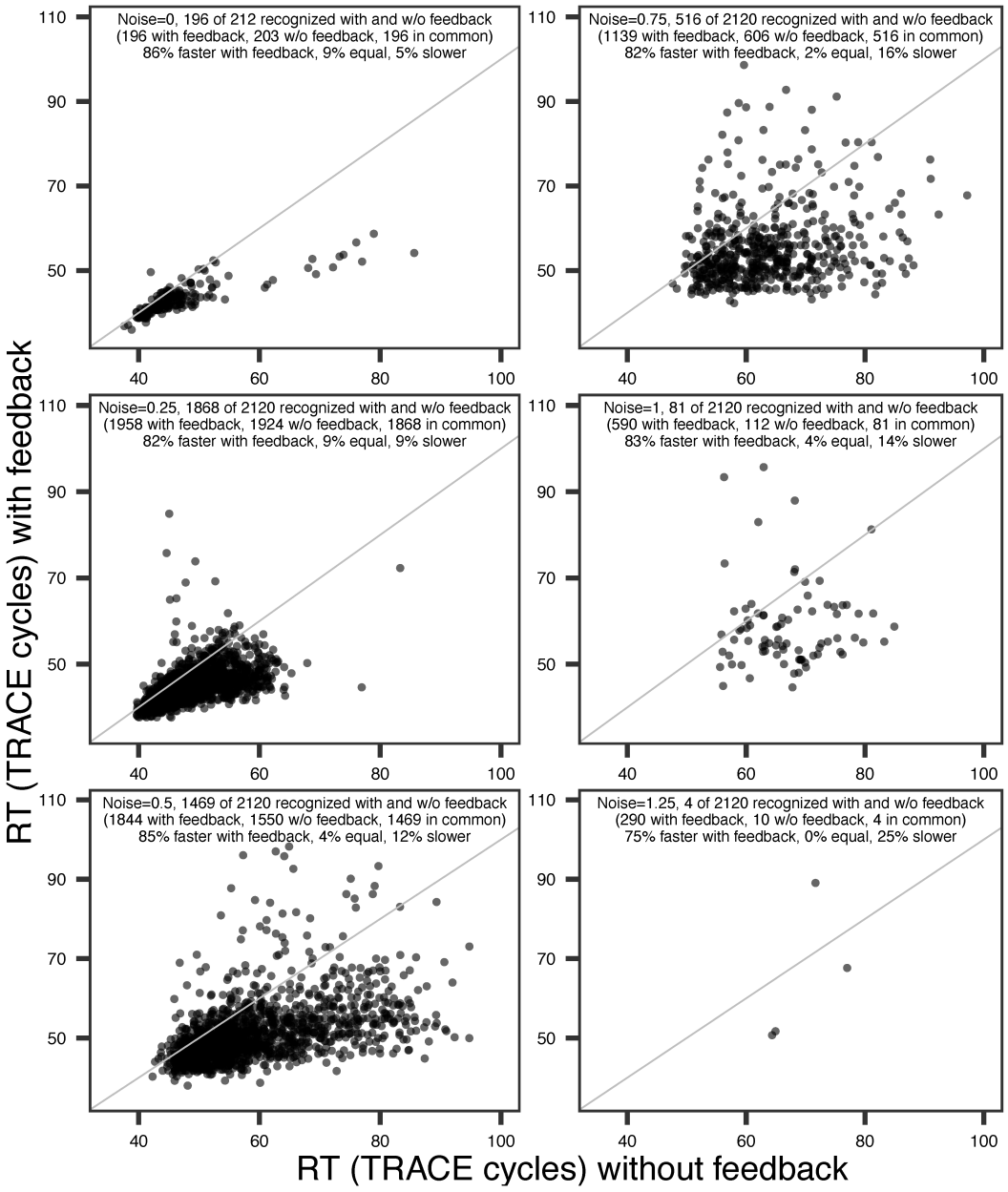

Figure A2. Comparing activation-based recognition time with feedback (set to 0.03) and without feedback (0.0) in Simulation 1. At each noise level greater than 0, there were 2120 simulations (10 repetitions of each word with Gaussian noise added to the input). Results are plotted only for words that were correctly recognized both with and without feedback. Items classified as "faster" were recognized more quickly (reached the threshold) more quickly with feedback than without; "equal" reached threshold at the same cycle with and without feedback; "slower" were reached the threshold later with feedback than without.