A blind spot in correct naming latency analyses

Gary M. Oppenheim
Bangor University, Bangor, Gwynedd, UK
Rice University, Houston, TX, USA

Abstract

Speech errors and naming latencies provide two complementary sets of behavioural data for understanding language production processes. A recent analytical trend – applied to intact and impaired production alike – highlights a link between specific features of correct picture naming latency distributions and the retrieval processes thought to underlie them. Although chronometric approaches to language production typically consider correct response times in isolation, adequately accounting for their distributions in error-prone situations requires also considering the errors that sometimes censor them. In this paper, I illustrate by simulation how excluding incorrect word retrievals predictably alters observed distributions of correct naming latencies. To the extent that naming errors impose a stochastic deadline on successful production, their censoring should tend to reduce the mean, variance, and skew of observed latencies for correct responses, relative to the uncensored underlying distribution.

Introduction

Imagine word retrieval as a series of races between a cat and a dog, where we're trying to discover what makes the cat run faster or slower, but can only record its winning times. We suspect that its fear of bulldogs might make it faster, and being intimidated by greyhounds might make it slower. But our races include two restrictions. First, each race lasts only 10 seconds at most; if the cat gets distracted, or injured, or quarrels with the dog, we can't record the cat's speed. Second, if the dog wins a race, the cat immediately gives up, so we can't record the cat's speed in this case, either. Thus recording only the cat's winning times filters its slower times from the dataset, changing our observed distributions. Racing against a (fast) greyhound, the cat will therefore seem exceptionally fast because we'll only record its times on those races where it outran a greyhound. Racing against a (slow) bulldog, the cat will seem much slower because we'll record its times in almost every race, even on its 'off' days. Although the cat may be spurred on by the bulldog, or discouraged by the greyhound – and these factors may even affect its distributions in unique ways – it will be difficult to distinguish their effects from the consequences of our recording biases.

Speech errors and naming latencies provide two complementary sources of behavioural data that any satisfying theory of language production must address. Because speakers produce both errors and correct responses in much the same way, errors can reveal processes, representations, and brain regions that normally subserve successful production; for instance, a *cat* → *rat* slip is more likely than a $cat \rightarrow \{mouse||mat||isotope||ngik\}$ slip, reflecting numerous semantic and phonological constraints. But because errors also – at least – reflect instances where production has somehow gone awry, it is useful to have converging evidence from measurements of correct productions: naming latencies. Naming latencies clearly reveal something about a word's accessibility: when producing a word is easier, speakers do it faster, revealing specific factors that facilitate and inhibit production. And because researchers typically restrict their naming latency analyses to correct productions – excluding any error trials – we often assume that latencies' means and distributions provide untainted windows onto successful production processes; this assumption is the focus of the current paper.

A simple difference in mean naming latencies is not as directly informative as the differences in probabilities *and* types of errors that form the traditional mainstay of neuropsychological investigations, but recent applications of distribution-fitting analyses (e.g. Anders et al., this volume; Roelofs & Piai, 2017; Scaltritti, Navarrete, & Peressotti, 2014) hold the promise of identifying specific aspects of decision and response processes that underlie such RT differences. The basic idea is that one can describe the shape of an RT distribution in terms of a small number of free parameters within a fixed mathematical formula. These are essentially data compression algorithms (in the same way that an *.MP3 or *.JPG is a *compressed* version of a larger computer file), but such distribution-fitting approaches are distinguished by an assumption that, because the algorithm involves fitting a plausible generative framework[1] to an observed distribution, the parameters that best describe the distribution have psychologically meaningful interpretations.[2] While increased attention to RT distributions is clearly a positive development, two major questions when interpreting distributions from an impaired system (or speaker) are how faithfully they reflect the same underlying processes as in an unimpaired system, and what kinds of systematic differences the impairment might introduce.

Although models of production have long addressed speech error (e.g. Dell, 1986; Fromkin, 1971) and naming latency (e.g. Levelt, Roelofs, & Meyer, 1999) data separately, recent accounts have begun integrating the two. For instance, Oppenheim, Dell, & Schwartz's (2010) Dark Side model combined an evidence-accumulation mechanism for lexical selection with an incremental learning process to demonstrate how the same implicit learning and unlearning of meaning-to-word associations could make unimpaired participants slower to name pictures correctly, and cause impaired participants to name them incorrectly or not at all. By quantitatively fitting a similar evidence-accumulation framework to humans' correct naming latency distributions (modeling just a single accumulator, without learning), Anders and colleagues' psychometric approach (2015; this volume) usefully identifies specific parameter changes that may underlie observed naming latency differences; a goal of process models like the Dark Side is to explain why some[3] such parameter changes might occur. Of course, addressing both errors and response times in separate studies or populations is an important first step, but truly integrating them requires better understanding their interplay within a single set of data.

An obvious example of this interplay can be found in omission cutoffs. Omission errors are operationally defined as failures to respond within a reasonable timeframe, which might be 2000ms for unimpaired participants, or 20,000ms for people with language impairments.[4] Because we define these outcomes in terms of their timecourse, classifying a slow would-be-correct response as an omission clearly affects our naming latency analyses: with a 2000ms deadline, a 1999ms *cat!* increases the observed latencies' mean and variance, while a 2001ms *cat!* – counted as an omission – does not.

So how do errors of commission affect latencies for correct responses? To address this question, we need a model of how such errors occur. The following minimal model allows us to focus on the interplay of naming errors (e.g. $cat \rightarrow \{dog||tiger||kitten\}$) and correct naming latencies, generalising to any reasonable account of production. This model simply assumes lognormal distributions of expected response times for target words and possible alternative names (or *competitors*) – with minimal consideration of their generating processes – and then considers how the observed RTs for correct responses change when alternatives stochastically emerge instead. Appendices A and B, respectively, demonstrate similar relationships when generating distributions via empirically derived parameters from fitted Shifted Wald and ex-Gaussian models.

Simulations

Starting with the assumption that a speaker semantically activates many candidate words, but ultimately selects just one for production, we can characterise word retrieval as a race between these candidates, where the most activated candidate will eventually win, and its retrieval time will depend, at a minimum[5], on the strength of its own activation (e.g. Oppenheim et al, 2010). The target word should usually be more activated than alternative names, but internal or external sources of variance can occasionally lead to it being underactivated and an alternative name overactivated, such that the alternative emerges instead; in this situation, we can consider such errors' effects on observed response time distributions for the target.

Formalising this approach, imagine that target word activations on 100,000 trials are normally distributed,

[1] Or an approximation thereof, in the case of ex-Gaussian analyses.

[2] In fact, the classic approach of comparing distributions in terms of means and variances is usually applied in this same way, assuming that differences in these parameters reflect particular psychological quantities.

[3] Whereas psychometric models typically collapse across processes and levels of representation to explain the outcome of an entire stimulus-to-response mapping, process models limit their scopes to consider specific processes and levels in more detail.

[4] Self-reported 'don't know' responses present a similar case, except that the deadline is stochastically self-imposed.

[5] For computational simplicity, this exercise implements noncompetitive selection, but all claimed patterns remain if assuming competition.

with a mean of 1 and a standard deviation of 0.4. For these same 100,000 trials, a single strong alternative tends to be slightly less activated, with a mean of 0.5 and a standard deviation of 0.4. Assume that the time it would take for any word to reach a selection threshold, considered individually, is simply the exponential of its negative activation (e^{-a}), approximating the duration of an evidence-accumulation decision process (cf. Heathcote & Love, 2012). Scaling this value by one constant and adding another ($1000 \times e^{-a} + 400$) adjusts it into a plausible millisecond response time. Considered independently, 100,000 trials cuing the target and 100,000 cuing the alternative predict the log-normal RT probability densities depicted in Figure 1a.

But because a person can only select one word on each trial, we assume that, on any trial where an alternative reaches threshold first, it preempts the target, censoring the target's RT from the observed distribution. By convention, we analyse only RTs for correct responses, so eliminating 18,767 (18.8%) error trials (where the alternative emerged instead) yields the modified probability density function shown in Figure 1b. Two remarkable changes emerge. First, because naming errors stochastically preempt slower correct responses, much like omission errors, they *reduce* mean RTs for the correct responses that remain.[6] Thus, within a single dataset, factors that promote errors may be counterintuitively associated with RT facilitation[7], producing apparent dissociations between error and RT effects despite a common source. Second, naming errors more generally reduce the variance of correct RTs, by reducing the distribution's positive skew. Table 1 confirms the intuition that such censoring affects multiple parameters when applying two common distribution-fitting analyses to the correct RTs. What is remarkable about this simulation is that these changes in correct RT distributions emerge without any underlying change in the target activation nor its retrieval process, which analyses restricted to correct RTs are commonly assumed to recover. Thus, correct RTs, in the presence of error effects, do not simply reflect the correct functioning underlying error-free production, but also reflect predictable biases from non-random data loss.

Similar biases emerge if correct responses are instead preempted by an array of ten weaker alternatives, with a mean activation of 0.01 and a standard deviation of 0.4 (Figures 1c and 1d). When comparing the target activation to only one randomly chosen alternative, there is little chance of an error (4%) and little effect on target RT distributions (but cf Shifted Wald parameters in Table 1). But considering all ten weak alternatives yields error rates (20.9%) and target RT distributions more similar to the previous 'strong alternative' simulation. Note, though, that the weak alternative array produces a different RT distribution than the single strong alternative, and the censored correct distribution reflects it. Thus, whereas the first simulation illustrated how errors affect target RT distributions by considering a single very strong alternative, the process naturally extends to situations where one has multiple weak (or strong) alternatives.

[6] Note that this differs from traditional conceptions of a speed/accuracy tradeoff, as it describes consequences of data censoring, rather than a response strategy.

[7] Or at least reduced interference – as the factors that promote alternative names often also weaken target responses, either directly or via competition.

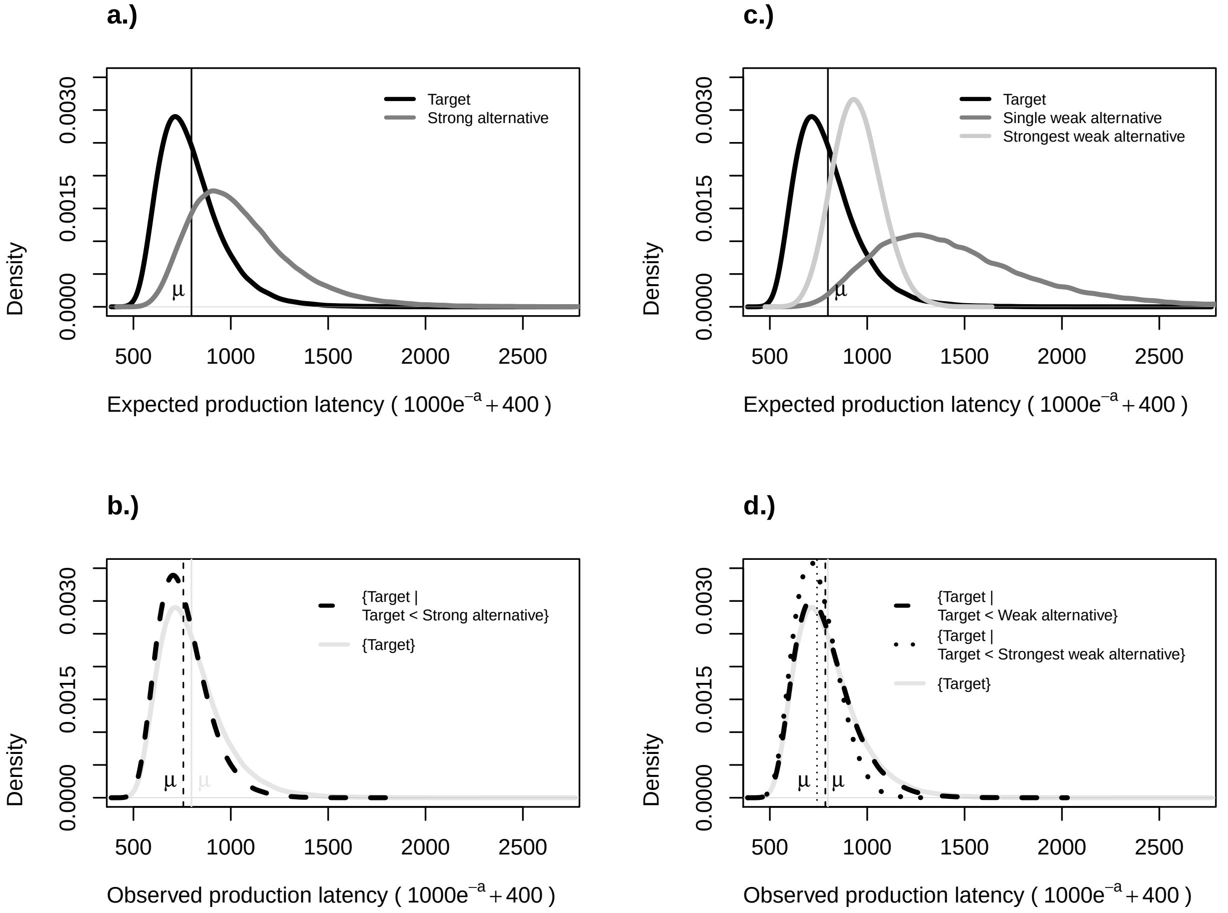

Figure 1. Densities of expected target and alternative selection times for 100,000 trials, before (a,c) and after (b,d) applying a single−output−per−trial constraint. Uncensored (a) and censored (b) densities for target selection in the presence of a strong alternative. Uncensored (c) and censored (d) densities for target selection in the presence of either one or ten weak alternatives.

Table 1: Accuracy, mean RT, and variance, plus Shifted Wald and ex-Gaussian fit parameters for the target RT distributions in Figure 1. In a fitted Shifted Wald model (via Heathcote's, 2004, fitwald()), α estimates the distance between a response's initial activation and its decision threshold, γ estimates the rate of evidence accumulation, and θ estimates time external to the accumulation process. In a fitted ex-Gaussian model (via Massidda's, 2013, retimes::mexgauss()), μ and σ estimate mean and variance, respectively, of a Gaussian component while τ estimates the skew from an exponential component. The same set of expected target RTs is used in all cases; the only differences are the properties of the competitor activations, and therefore their likelihood of censoring the distribution of correct response times.

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\mathrm{RT}_{\mathrm{Target}}$	1.00	798	27855	50	0.12	387	652	80	146
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Weak\ competitor}}$	0.96	785	22503	58	0.14	362	664	89	121
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Strong\ competitor}}$	0.81	756	16304	66	0.16	342	657	81	99

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\text{RT}_{\text{Target}} < \text{RT}_{\text{Fastest weak competitor}}$	0.79	742	12099	104	0.20	233	671	84	71

Discussion

To consider how error-based censoring affects distributions of correct RTs, here I manipulated word error rates by changing the accessibility of alternative names while holding target accessibility constant. Although these simulations are quite basic, they importantly integrate error production with the analysis of correct naming latencies to delineate an expected consequence of the former on the latter: simply because correct responses must outpace possible errors, *ceteris paribus* any factor that promotes premptive errors will also hasten mean *correct* word retrieval times by, essentially, imposing a stochastic limit on how slow any response can be before another preempts it. As errors grow more frequent, observed target RT distributions shift earlier, decreasing their means and variances[8] and reducing a positive skew in their distributions. Within an analytical framework that interprets correct RT distributions in terms of specific cognitive processes, without considering errors, such distribution changes would erroneously indicate changes in multiple psychologically interpretable parameters. More generally, these results challenge a popular assumption that RTs for correct responses provide a pristine window onto error-free production processes (e.g. Levelt et al, 1999).

Although specifically considering word selection errors, these conclusions should at a minimum generalise to any non-target outcome driven by difficulty accessing a target word (including, e.g., hesitations, 'acceptable' alternatives), and generalise to analogous conflicts at other levels of representation and in other response domains. It can be easily demonstrated that more developed generative models (e.g. Oppenheim et al.'s 2010 Dark Side model) exhibit similar behaviors, but here I have used a simpler mathematical approach to emphasize that these general patterns should emerge from any account where the speed of target retrieval and the probability of an error depend in some way on the activation of the target response, which is to say any reasonable theory of word retrieval. Although the extent and type of changes to an observed target distribution depend on characteristics of the censoring distribution, the only cases that would fully escape such effects are those where either target productions fail completely at random (without regard to their strength), or target responses have been completely lost (thus for patient work it would be useful to distinguish temporary paraphasias from wholesale word deficits).

The informative censoring considered here is far from the only factor that affects target RT distributions, and error rates can of course also be affected by changes in the accessibility of targets as well as alternatives. For instance, the Dark Side model's account of cumulative semantic interference proposes that semantic errors reflect target weakening as well as 'competitor' strengthening, and under competitive selection anything promoting an alternative should consequently weaken the target. Thus, in practice we may expect to find error effects paired with RT effects that are merely less robust than expected, rather than fully complementary (but see Mahon, Costa, Peterson, Vargas, & Caramazza, 2007, for a prominent empirical example of such a dissociations in the context of picture-word interference).

Given this interplay between error and response time effects – and a lack of robust statistical techniques to correct for non-random missing data[9] – how can we interpret response time data in the presence of error effects? I see three possibilities, ordered from simplest to most complex. The first – and most frequent in neuropsychology – entirely avoids analysing RTs in the presence of error effects (or error rate differences, in between-subject comparisons). Avoidance makes some sense, also considering that one might generally expect RTs that accompany high error rates to be dominated by noise, and thus unlikely to reveal reliable patterns. In the interest of using every part of the buffalo, a second approach might allow RT analyses,

[8] This linked decrease is consistent with observations that these statistics correlate empirically (e.g. Wagenmakers & Brown, 2007).

[9] Quantitative correction remains an active area of research (e.g. Enders, 2010), but because proposed methods tend to correct by introducing assumption-based compensatory biases, they necessarily introduce incorrect biases when the assumptions are imperfect. Moreover, such corrections would still only allow approximations of data in the long tail of a distribution, where distribution-fitting approaches tend to be rather volatile.

while acknowledging possible biases introduced by the error patterns. Thus, finding complementary or incommensurate error and RT effects within the same participants would not necessarily imply that the effects reflect distinct underlying processes (as might seem reasonable to conclude, a priori). But an overpowered and underconstrained caveat introduces the risk of explaining-away valid data. Therefore, a third approach – and a far better way of using the buffalo – is to quantify the expected biases, generating predicted error and RT data by explicitly simulating the underlying processes; such integrated approaches have dealt successfully with simpler decisions in other domains (e.g. Ratcliff, 1978). Quantitatively fitting observed distributions for error-prone picture naming, an *n*-alternative quasi-forced-choice task, may prove infeasible – requiring fitting individual distributions for each alternatives decision latencies – but even an approximation should usefully constrain predictions of such interdependencies. Although the Dark Side model's oversimplified information accumulation process was not initially developed for the task, reasonable extensions (Oppenheim, in prep) may allow it to usefully serve in this capacity.

Conclusion

Although errors and response times both provide important clues about language production processes, we typically consider them independently. Re-integrating these approaches requires considering not only how speed affects accuracy, but also how accuracy affects (observed) speed. This step is particularly important when considering correct response times in the presence of frequent errors: target retrieval processes, considered alone, are a useful place to start, but may not adequately explain observed target retrieval times. Error productions affect target RT distributions, limiting our ability to meaningfully interpret such distributions in terms of underlying cognitive processes. However, to the extent that errors affect distributions predictably, process models that integrate errors and response times may usefully estimate the direction and magnitude of any biases that they introduce.

References

Anders, R., Ries, S., van Maanen, L., & Alario, F. X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology, 82,* 57–73.

Anders, R., Ries, S., Van Maanen, L., & Alario, F.-X. (in press). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology.*

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93(3),* 283–321.

Enders, C. K. (2010). *Applied Missing Data Analysis.* Guilford Press.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47(1),* 27–52.

Heathcote, A. (2004). Fitting wald and ex-Wald distributions to response time data: an example using functions for the S-PLUS package. *Behavior Research Methods, Instruments, & Computers, 36(4),* 678–694.

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3(AUG),* 1–19.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22(1),* 1–38.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology-Learning Memory and Cognition, 33(3),* 503–535.

Massidda, D. (2013). retimes: Reaction time analysis [Software] (R package version 0.1-2). Retrieved from http://CRAN.R-project.org/package=retimes

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition, 114(2),* 227–252.

Roelofs, A., & Piai, V. (2017). Distributional analysis of semantic interference in picture naming. *The Quarterly Journal of Experimental Psychology, 70(4),* 782–792.

Scaltritti, M., Navarrete, E., & Peressotti, F. (2014). Distributional analyses in the picture-word interference paradigm: Exploring the semantic interference and the distractor frequency effects. Quarterly Journal of Experimental Psychology (2006), 218(December 2014), 1–22. http://doi.org/10.1080/17470218.2014.981196

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114(3),* 830–841.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85(2),* 59–108.

Appendix A

Generating distibutions via a Shifted Wald function

Simulations

Starting with the assumption that a speaker semantically activates many candidate words, but ultimately selects just one for production, we can characterise word retrieval as a race between these candidates, where the most activated candidate will eventually win, and its retrieval time will depend, at a minimum[10], on the strength of its own activation (e.g. Oppenheim et al, 2010). The target word should usually be more activated than alternative names, but internal or external sources of variance can occasionally lead to it being underactivated and an alternative name overactivated, such that the alternative emerges instead; in this situation, we can consider such errors' effects on observed response time distributions for the target.

Formalising this approach, imagine that target word RTs on 100,000 trials are generated from a random walk process, implemented as sampling from a Shifted Wald distribution with parameters $\{\alpha = 30,\ \gamma = 0.15,\ \theta = 565\}$ (approximating parameters in Anders et al., this volume), where α represents the distance between the word's initial activation and its selection threshold, γ represents the mean rate of evidence accumulation (i.e. the mean slope of the random walk), and θ represents the time taken by all other aspects of the stimulus→response process (both before and after the decision process). For these same 100,000 trials, a single strong alternative tends to be slightly less activated; in a Shifted Wald model, less activation might map onto the idea that dispreferred responses have higher α ('threshold') values, indicating greater distance between the starting point and the threshold, and lower γ ('accumulation rate') values, yielding the parameters $\{\alpha = 40,\ \gamma = 0.1,\ \theta = 565\}$. Considered independently, 100,000 trials cuing the target and 100,000 cuing the alternative predict the RT probability densities depicted in Figure A1a.

But because a person can only select one word on each trial, we assume that, on any trial where an alternative reaches threshold first, it preempts the target, censoring the target's RT from the observed distribution. By convention, we analyse only RTs for correct responses, so eliminating 14,972 (15%) error trials (where the alternative emerged instead) yields the modified probability density function shown in Figure A1b. Two remarkable changes emerge. First, because naming errors stochastically preempt slower correct responses, much like omission errors, they *reduce* mean RTs for the correct responses that remain.[11] Thus, within a single dataset, factors that promote errors may be counterintuitively associated with RT facilitation[12], producing apparent dissociations between error and RT effects despite a common source. Second, naming errors more generally reduce the variance of correct RTs, by reducing the distribution's positive skew. Table 2 confirms the intuition that such censoring affects multiple parameters when applying two common distribution-fitting analyses to the correct RTs. What is remarkable about this simulation is that these changes in correct RT distributions emerge without any underlying change in the target activation nor its retrieval process, which analyses restricted to correct RTs are commonly assumed to recover. Thus, correct RTs, in the presence of error effects, do not simply reflect the correct functioning underlying error-free production, but also reflect predictable biases from non-random data loss.

Similar biases emerge if correct responses are instead preempted by an array of ten weaker alternatives, generated from ex-Gaussian parameters $\{\alpha = 50,\ \gamma = 0.05,\ \theta = 565\}$ (Figures A1c and A1d). When comparing the target activation to only one randomly chosen alternative, there is little chance of an error (1.7%) and little effect on target RT distributions (but cf Shifted Wald parameters in Table 2). But considering all ten weak alternatives yields error rates (11.9%) and target RT distributions a bit more similar to the previous 'strong alternative' simulation. Note, though, that the weak alternative array produces a different RT distribution than the single strong alternative, and the censored correct distribution reflects it. Thus, whereas the first

[10] For computational simplicity, this exercise implements noncompetitive selection, but all claimed patterns remain if assuming competition.

[11] Note that this differs from traditional conceptions of a speed/accuracy tradeoff, as it describes consequences of data censoring, rather than a response strategy.

[12] Or at least reduced interference – as the factors that promote alternatives often also weaken target responses, either directly or via competition.

simulation illustrated how errors affect target RT distributions by considering a single very strong alternative, the process naturally extends to situations where one has multiple weak (or strong) alternatives.

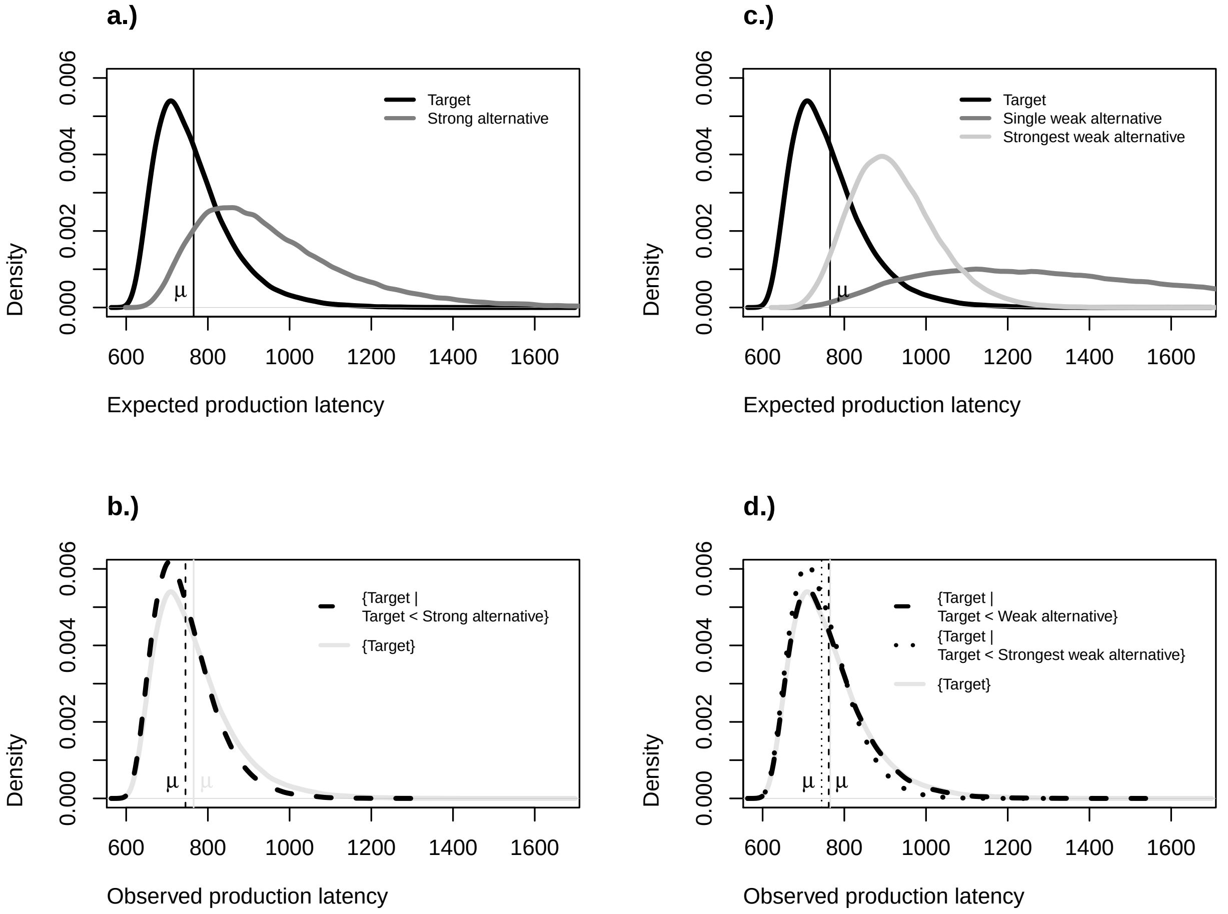

Figure 1. Densities of expected target and alternative selection times for 100,000 trials, before (a,c) and after (b,d) applying a single−output−per−trial constraint. Uncensored (a) and censored (b) densities for target selection in the presence of a strong alternative. Uncensored (c) and censored (d) densities for target selection in the presence of either one or ten weak alternatives.

Table 2: Accuracy, mean RT, and variance, plus Shifted Wald and ex-Gaussian fit parameters for the target RT distributions in Figure A1. In a fitted Shifted Wald model (via Heathcote's, 2004, fitwald()), α estimates the distance between a response's initial activation and its decision threshold, γ estimates the rate of evidence accumulation, and θ estimates time external to the accumulation process. In a fitted ex-Gaussian model (via Massidda's, 2013, retimes::mexgauss()), μ and σ estimate mean and variance, respectively, of a Gaussian component while τ estimates the skew from an exponential component. The same set of expected target RTs is used in all cases; the only differences are the properties of the competitor activations, and therefore their likelihood of censoring the distribution of correct response times.

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\mathrm{RT}_{\mathrm{Target}}$	1.00	765	8906	30	0.15	567	681	43	84

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Weak\ competitor}}$	0.98	762	8127	31	0.16	564	683	45	78
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Strong\ competitor}}$	0.85	745	5684	34	0.18	557	682	42	63
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Fastest\ weak\ competitor}}$	0.88	744	5039	40	0.20	542	690	46	54

Appendix B

Generating distibutions via an Ex-Gaussian function

Starting with the assumption that a speaker semantically activates many candidate words, but ultimately selects just one for production, we can characterise word retrieval as a race between these candidates, where the most activated candidate will eventually win, and its retrieval time will depend, at a minimum[13], on the strength of its own activation (e.g. Oppenheim et al, 2010). The target word should usually be more activated than alternative names, but internal or external sources of variance can occasionally lead to it being underactivated and an alternative name overactivated, such that the alternative emerges instead; in this situation, we can consider such errors' effects on observed response time distributions for the target.

Formalising this approach, imagine that target word RTs on 100,000 trials are sampled[14] from an ex-Gaussian distribution, with parameters $\{\mu = 619, \sigma = 45, \tau = 123\}$ (parameters taken from Roelofs & Piai's, 2017, empirical estimations from a neutral distractor condition in picture-word interference). For these same 100,000 trials, a single strong competitor tends to be slightly less activated, implemented here as ~25% increases in all parameters, $\{\mu = 800, \sigma = 60, \tau = 200\}$. Considered independently, 100,000 trials cuing the target and 100,000 cuing the alternative predict the log-normal RT probability densities depicted in Figure B1a.

But because a person can only select one word on each trial, we assume that, on any trial where an alternative reaches threshold first, it preempts the target, censoring the target's RT from the observed distribution. By convention, we analyse only RTs for correct responses, so eliminating 10,536 (10.5%) error trials (where the alternative emerged instead) yields the modified probability density function shown in Figure B1b. Two remarkable changes emerge. First, because naming errors stochastically preempt slower correct responses, much like omission errors, they *reduce* mean RTs for the correct responses that remain.[15] Thus, within a single dataset, factors that promote errors may be counterintuitively associated with RT facilitation[16], producing apparent dissociations between error and RT effects despite a common source. Second, naming errors more generally reduce the variance of correct RTs, by reducing the distribution's positive skew. Table 3 confirms the intuition that such censoring affects multiple parameters when applying two common distribution-fitting analyses to the correct RTs. Notably, although the error-based censoring primarily reduces the long right tail of the correct RT distribution, thus decreasing τ, here it also *increases* μ to compensate, illustrating the parameters interdependency. What is remarkable about this simulation is that these changes in correct RT distributions emerge without any underlying change in the target activation nor its retrieval process, which analyses restricted to correct RTs are commonly assumed to recover. Thus, correct RTs, in the presence of error effects, do not simply reflect the correct functioning underlying error-free production, but also reflect predictable biases from non-random data loss.

Similar biases emerge if correct responses are instead preempted by an array of ten weaker alternatives, generated from ex-Gaussian parameters $\{\mu = 1000, \sigma = 75, \tau = 250\}$ (Figures B1c and B1d). When comparing the target activation to only one randomly chosen alternative, there is little chance of an error (2.0%) and little effect on target RT distributions (but cf fitted model parameters in Table 3). But considering all ten weak alternatives yields error rates (6.9%) and target RT distributions a bit more similar to the previous 'strong alternative' simulation. Note, though, that the weak alternative array produces a different RT distribution than the single strong alternative, and the censored correct distribution reflects it. Thus, whereas the first simulation illustrated how errors affect target RT distributions by considering a single very strong alternative, the process naturally extends to situations where one has multiple weak (or strong) alternatives.

[13] For computational simplicity, this exercise implements noncompetitive selection, but all claimed patterns remain if assuming competition.

[14] I use the term 'sampling' here because, although researchers often identify parameter changes with particular psychological processes, ex-Gaussian analyses tend not to assume that RTs are actually generated by an ex-Gaussian mental process.

[15]

[16]

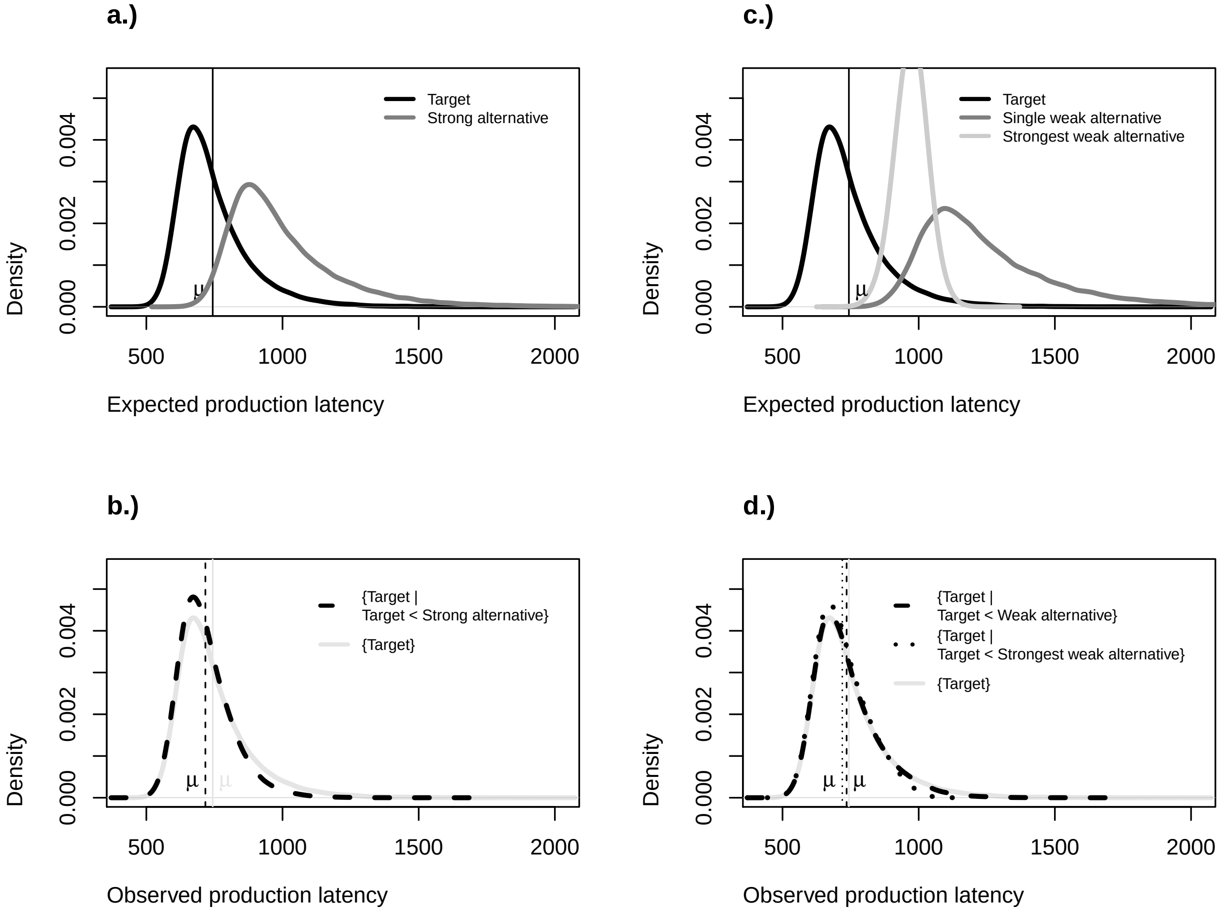

Figure 1. Densities of expected target and alternative selection times for 100,000 trials, before (a,c) and after (b,d) applying a single−output−per−trial constraint. Uncensored (a) and censored (b) densities for target selection in the presence of a strong alternative. Uncensored (c) and censored (d) densities for target selection in the presence of either one or ten weak alternatives.

Table 3: Accuracy, mean RT, and variance, plus Shifted Wald and ex-Gaussian fit parameters for the target RT distributions in Figure B1. In a fitted Shifted Wald model (via Heathcote's, 2004, fitwald()), α estimates the distance between a response's initial activation and its decision threshold, γ estimates the rate of evidence accumulation, and θ estimates time external to the accumulation process. In a fitted ex-Gaussian model (via Massidda's, 2013, retimes::mexgauss()), μ and σ estimate mean and variance, respectively, of a Gaussian component while τ estimates the skew from an exponential component. The same set of expected target RTs is used in all cases; the only differences are the properties of the competitor activations, and therefore their likelihood of censoring the distribution of correct response times.

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\mathrm{RT}_{\mathrm{Target}}$	1.00	743	17433	41	0.14	443	618	41	125
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Weak\ competitor}}$	0.98	735	13456	48	0.15	426	635	60	99
$\mathrm{RT}_{\mathrm{Target}} < \mathrm{RT}_{\mathrm{Strong\ competitor}}$	0.89	716	9427	59	0.19	401	636	55	80

	Accuracy	Mean RT	Variance	Shifted Wald			Ex-Gaussian		
				α	γ	θ	μ	σ	τ
$\text{RT}_{\text{Target}} < \text{RT}_{\text{Fastest weak competitor}}$	0.93	719	8734	72	0.20	360	655	68	64