

*Should be cited as:* Pinet, S., & Nozari, N. (2021). The role of visual feedback in detecting and correcting typing errors: A signal detection approach. *Journal of Memory and Language*, 117, 104193. <http://doi.org/10.1016/j.jml.2020.104193>

## The role of visual feedback in detecting and correcting typing errors:

### A signal detection approach

Svetlana Pinet<sup>1,2</sup> and Nazbanou Nozari<sup>3,4</sup>

1. BCBL. Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, Spain

2. Department of Neurology, Johns Hopkins University, Baltimore, USA

3. Department of Psychology, Carnegie Mellon University, Pittsburgh, USA

4. Center for Neural Basis of Cognition (CNBC)

Correspondence may be addressed to:

Svetlana Pinet (corresponding author)

Address: Paseo Mikeletegi 69, 2nd Floor

20009 Donostia/San Sebastian, Spain

Email: s.pinet@bcbl.eu

Nazbanou Nozari, MD, PhD

Address: 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Email: bnozari@andrew.cmu.edu

## **Abstract**

This study examined the role of external information in monitoring language production. In a typing-to-dictation task, participants were deprived of all or part of visual feedback. Data were analyzed using signal detection theory (SDT) applied to a multi-component monitoring framework. Results showed that removing the visual information affected the correction of typing errors more than their conscious detection (Exps 1, 2). Reinstating partial visual information (positional information) increased correction rates but not to the level of full visual information, independently of the probability of error detection (Exps 2, 3). Analysis of SDT parameters showed that while manipulating visual information affected the informativeness of the signal for both correction and conscious detection of errors, participants treated this change differently in the two tasks. We discuss the implications of the results, and more generally, the utility of SDT, for theories of monitoring and control in language production.

## **Highlights**

- Three experiments manipulating visual feedback while typing single words
- Error detection moderately impacted by visual feedback removal
- Error correction relying heavily on visual feedback
- Positional information helpful, but letter identity essential for corrections
- Application of signal detection theory to multichannel language monitoring

## **Keywords**

Language production, signal detection theory, performance monitoring, typing, error repair

## **Introduction**

Performance monitoring refers to the processes involved in the surveillance of one's cognitive and motor operations in order to ensure a satisfactory outcome. It is thus true, by definition, that monitoring must be sensitive to the output of cognitive processes; we realize that we have spoken the wrong word when we hear ourselves, or realize that a typing error has occurred when the incorrect letter appears on the screen. In fact, sensory processing of the output (i.e., "external monitoring") can probably provide the most reliable feedback on one's own performance, as there is no uncertainty about the outcome once the output is externally available. The downside of external monitoring is that it comes too late, limiting the scope of monitoring to detecting errors that have already been made, as opposed to preventing errors from happening in real time. To solve this problem, "internal monitoring" mechanisms have been proposed, which, despite their various flavors, share the premise that monitoring need not rely on the final output of the process. An important question is how the labor is divided between internal and external mechanisms. Obviously, in the learning phase of a task, when performance is very error-prone and strong internal representations are not yet available, adjustments must be guided by performance outcome. Thus, learning is expected to depend heavily on the external monitoring. Less well known is the role of external monitoring once a task is learned and internal representations are well developed. This study addresses this question.

### **Performance monitoring in language production**

Language production provides an excellent testbed for performance monitoring in real life for several reasons: (1) it is a complicated generative task that captures the true demands on the cognitive system better than artificial two-choice button-press tasks. (2) It is an extremely well-practiced task, which obviates the need for training in the lab and consequently avoids differences in how well people have learned the artificial task. (3) Avoiding errors and maintaining fluency are two natural and simultaneous goals of language production that people have years of experience balancing. For these reasons, studying monitoring in language production is not only informative about the specific case of language, but also more broadly of how humans monitor complex cognitive behavior.

Monitoring models in language production include both external (e.g., external channel in the perceptual loop; Levelt, 1983; DIVA, Guenther, 1995; Guenther, 2016; Tourville & Guenther, 2011), and internal monitoring mechanisms (e.g., internal channel of the perceptual loop, Hartsuiker & Kolk, 2001; Levelt, 1983; or other mechanisms, e.g., Hickok, 2012; Nooteboom & Quené, 2019; Nozari, Dell, & Schwartz, 2011; Pickering & Garrod, 2013; see Nozari & Novick, 2017, and Nozari, 2020, for reviews). The label “external channel” in this literature refers to mechanisms that process the final outcome of production, once it becomes available externally for processing through the perceptual system. “Internal channel”, on the other hand, refers to processes that rely heavily on internal representations with or without the involvement of the perceptual system. External monitoring mechanisms are particularly powerful for explaining the fine-tuning of a system which aims to produce a nuanced output, such as children mastering the native phonology of their language (Guenther, 2016). They, however, have trouble accounting to very quick repairs. For instance, in the classic example “v- horizontal” from Levelt (1983), the repair starts immediately after the interruption of the first segment of the word, with no delay. This rapid timeline is incompatible with an external monitor, which would need to process the acoustic signal, detect its mismatch to the intended signal, issue a command to halt ongoing behavior and plan a repair. Internal monitoring models account for such findings by proposing mechanisms that can detect an error internally, i.e., before it is articulated. Extensions of some of those models also naturally explain behaviors beyond error detection, for example why speakers take longer to produce words when the chance of making an error is higher (Nozari & Hepner, 2019a, 2019b).

The question of the division of labor between external and internal monitoring channels has been investigated in spoken production, mostly by contrasting error detection under circumstances when auditory feedback is available or blocked by noise (e.g., Nooteboom & Quené, 2017; Oomen & Postma, 2001; Postma & Kolk, 1992; Postma & Noordanus, 1996). In a classic study of auditory blocking, Lackner and Tuller (1979) had participants push a button if they detected their speech errors in conditions with and without noise. The results showed that participants not only frequently detected their errors in the noise-masked condition, but that they were also faster to do so compared to the no-noise condition. These findings were taken as evidence for fast and efficient monitoring through the internal channel without any contribution from the external channel (see Postma & Kolk, 1992 for similar findings). Converging evidence

from other studies also points to a quick and efficient internal monitoring channel (Hartsuiker, Pickering, & de Jong, 2005). For instance, fast interruptions (i.e., short error-to-cut-off intervals) and repairs (i.e., short cut-off-to-repair intervals) indicate that the response must have been monitored and repair processes may have been already initiated before articulation began (Postma & Oomen, 2005).

The evidence reviewed above points to an efficient internal channel, capable of error monitoring even in the absence of the external channel. The question then is: Does the external channel have any influence on monitoring errors in language production? Two lines of evidence suggest that it does. The first was provided by a clever study by Lind, Hall, Breidegard, Balkenius, and Johansson (2014) in which the authors manipulated auditory feedback such that on a small subset of trials, participants heard a different word from what they had produced. For example, they heard “green” instead of “grey”, while they were performing a Stroop task in which they had to name the ink color of a color word. When the lag between production and playback was short, participants reported the artificially inserted word as their own production on 68% of trials, thereby accepting an error they had not produced based on information from the external channel. These results demonstrate convincingly that when the external and internal monitoring systems process different information, the former can sometimes overwrite the latter. This situation, however, is not common in everyday life.

The second line of evidence in favor of some contribution of external monitoring in language production comes from studies showing a decrease in the proportion of detected and corrected errors under noise-masked conditions. For example, Postma and Noordanus (1996) had participants explicitly report their errors by pressing a button while reciting tongue-twisters and found a slightly lower rate of reporting errors under noise-masked than normal conditions (50% vs. 64%, respectively). Other studies have used correction performance as a proxy for monitoring. For example, Postma & Kolk (1992) had participants recite tongue-twisters and produce natural sentences in conditions with and without noise. While error rates were comparable in the two conditions, correction rates were about 50% lower under noise for both tongue-twisters and sentences. When accuracy was emphasized, participants made fewer errors and corrected a higher proportion of those errors, but the difference in the repair rate between conditions with and without noise persisted. Oomen et al. (2001) also found lower correction

rates under noise (a 17% decrease). More recently, however, Nooteboom & Quené (2017) found no difference in correction rates with and without noise-masking.

To summarize, the evidence in favor of internal monitoring in spoken production is convincing. The existence of an external channel is also undisputed. More controversial is the contribution of the external channel to monitoring in everyday production. It would also be desirable to differentiate between the role of external channel in conscious awareness of errors vs. applying corrections. Below, we discuss why typing provides a better medium than spoken production for addressing these issues.

## **Monitoring in typing**

As discussed in the previous section, one of the best methods for assessing the contribution of the external channel to monitoring is to remove its contribution, and examine the change to indices of monitoring such as conscious detection and correction of errors when the only available channel is the internal channel. In practice, this approach runs into non-trivial problems when applied to spoken production. For one thing, noise masking does not completely block the external channel because of bone conduction. For external noise to override any trace of self-produced auditory input, it has to be very loud, in which case the distraction and disturbance that it causes the speaker makes it difficult to claim that it only differs from the control condition in that it has removed the contribution of the external channel. Moreover, blocking the auditory input activates the Lombard reflex which causes changes to the primary production processes such as fundamental frequency and speech rate (Lane & Tranel, 1971), again making this condition different from the control condition in more than just one way. Even the ideal designs cannot overcome these physiologically induced constraints.

Other modalities of language production, however, are not subject to similar constraints. When the output is visual, as in typing, it can be completely removed without causing additional disturbance or distraction. Typing could thus provide a strong alternative for investigating the contribution of the external channel to performance monitoring. Typing has a special place in cognition as an activity that is situated at the intersection between language, motor, and visual processing. Its strong similarity to spoken language production has been demonstrated by showing the influence of psycholinguistic factors on typed performance indices, as well as the common error types in typing (Pinet & Nozari, 2018; Pinet, Ziegler, & Alario, 2016). Moreover, our past work has shown that principles such as feedback from the sublexical to the lexical layer that have been established in spoken production (e.g., Dell, 1984), are also evident from the

pattern of typing errors, even after the potential contribution of phonological processing is partialled out (Pinet & Nozari, 2018).

More recently, we have also demonstrated that monitoring in typing has the same electrophysiological signatures previously reported in non-linguistic forced-choice tasks, such as the error-related negativity (ERN), error-related positivity (Pe), feedback-related negativity (FRN), and frontocentral positivity (FCP; Pinet & Nozari, 2020). Of these, ERN has also been recovered in spoken production (e.g., Riès, Janssen, Dufau, Alario, & Burle, 2011), further supporting the theoretical position that typing and spoken production share similar domain-general principles of monitoring with each other and with other cognitive tasks (e.g., Hanley, Cortis, Budd, & Nozari, 2016; Nozari et al., 2011; see Roelofs, 2020 for a criticism and Nozari, 2020, for a response to this criticism). These similarities, together with the absence of the problems associated with blocking the external channel, make typing a good candidate for studying the contribution of the external channel to monitoring in language production and more broadly for complex cognitive tasks. The claim, of course, is not that the findings can be immediately extrapolated to spoken production, as there are inarguable differences between speaking and typing, such as amount of practice and the motor processes involved. Instead, these findings suggest that typing can provide clean information about the division of labor between two monitoring channels in a task that similarly taps into domain-general monitoring mechanisms.

To our knowledge, four studies have, to date, investigated the role of the external channel in monitoring in typing. Logan & Crump's (2010) manipulation was identical to Lind et al.'s (2014) study. They manipulated the typing output on the screen to either insert (on 6% of trials) or correct errors (on ~45% of trials). Expert typists accepted the inserted errors as their own on around 70% of such trials and took credit for the inserted corrections on around 90% of the artificially corrected trials. Interestingly, when they were briefed that manipulations were possible, and were given the additional options of “inserted” or “corrected” errors, they were able to detect 80% of inserted and 50% of corrected errors as such. This latter finding confirms that the external channel can indeed overwrite the internal channel, but that the extent of its contribution may be overestimated in studies that artificially manipulate the external output.

Three other studies have used the technique of removing the contribution of the external channel by preventing the immediate appearance of the visual output. These studies used participants that have a high level of proficiency in typing, but who were not necessarily formally trained and were therefore a better representative of the general population. Kalfaoglu & Stafford (2014) used a sentence-copying task without visual feedback and found that

participants were able to correct 60% of their errors in the absence of any visual information. This study, however, did not include a control condition with real-time visual feedback. Snyder, Logan, & Yamaguchi (2015) used a word copying task with and without real-time visual feedback and found lower rates of error detection (29% decrease) without visual feedback, as measured by explicit reports. The study, however, did not allow participants to correct their errors. Moreover, copy-typing encourages reliance on the visual modality that may be absent in typing words from meaning.

Most recently, in an EEG study, we had participants type single words that they heard with immediate and delayed visual feedback, and were allowed to use the backspace to correct their errors (Pinet & Nozari, 2020). Before displaying the visual output in the delayed-feedback condition, participants were explicitly asked if they had made an error. Performance was compared between the two feedback conditions. While accuracy and response times (RTs) were comparable between the two conditions, inter-keystroke intervals (IKIs) were significantly longer in the condition without immediate visual feedback. Importantly, we observed a steep drop in the correction rate (from 42% to 9%). Explicit judgments in the delayed-feedback condition revealed that participants were aware of 54% of committed errors in this condition, suggesting that the absence of corrections could not be simply attributed to the fact that errors had not been detected. The results of this study suggested considerable reliance on external feedback for error correction, higher than reported in any previous studies. However, the sample size was relatively small (17 participants). Moreover, the study did not include explicit metacognitive judgments on error awareness in the immediate-feedback condition, thus we were unable to determine whether the external channel had any role in conscious detection of errors or not. Finally, while the results suggested a significant dependence of repairs on visual feedback, the study provided no information on *why* that was the case, i.e., what kind of information in the visual output was deemed necessary for attempting repairs? The current study was designed to answer these questions.

## **Current study**

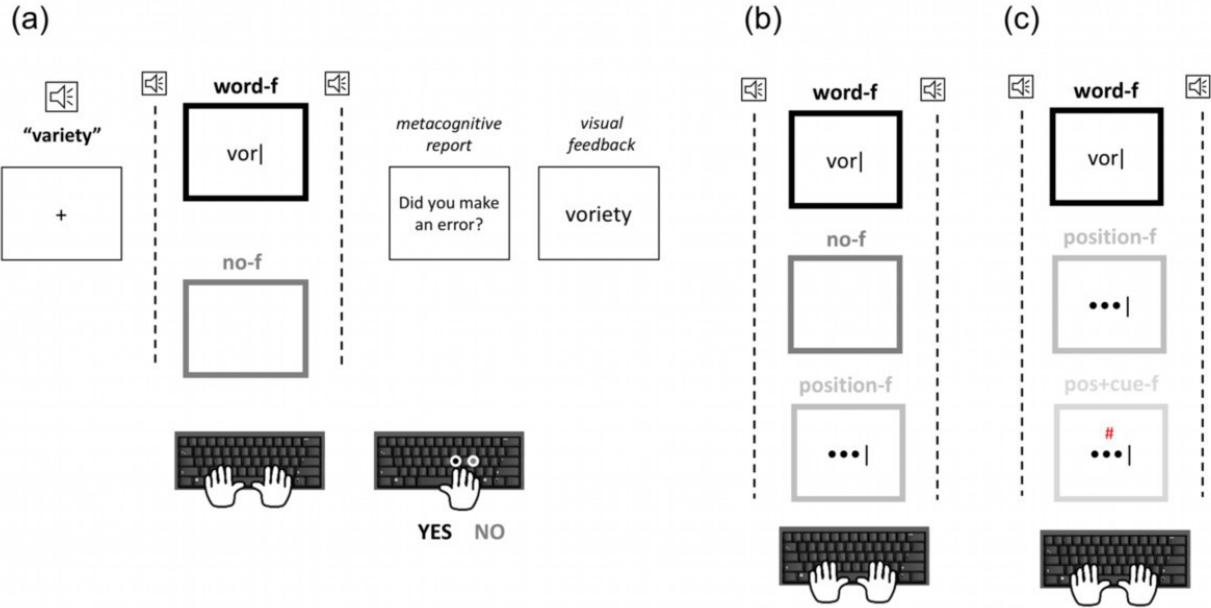
In three experiments, we expanded the design of Pinet & Nozari (2020) to investigate the contribution of the external channel, i.e., visual output processing, to conscious detection and correction of errors. All experiments use a typing-to-dictation paradigm, in which participants

type a real English word upon hearing it. Unlike copy-typing (i.e., typing by looking at the written forms of words), typing-to-dictation requires retrieving the lexical representation, as English orthography is not transparent (e.g., participants cannot type “yacht” correctly unless they first map sounds on to the proper lexical item, access its spelling, and then retrieve its segments). The production part of the task, thus, taps into the second stage of word production, i.e., mapping lexical items to segments (see Dell, Nozari, & Oppenheim, 2014 for a detailed review of the stages of word production). This claim is supported by the evidence showing that the typing-to-dictation performance is sensitive to lexical frequency (Bonin, Méot, Lagarrigue, & Roux, 2015; Pinet et al., 2016) which has a strong locus in mapping lexical items to segments (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008; Nozari, Kittredge, Dell, & Schwartz, 2010), as well as the properties of errors in this task which point to sequencing problems typical of mapping lexical items to segments (e.g., Hepner, Pinet, & Nozari, 2018). This makes the task ideal for studying segmental errors in language production.

In all experiments, participants heard a word and typed it before a deadline with the option of correcting their errors as in real life. Once they finished typing the word, they were asked to judge whether they had made an error or not (the metacognitive judgment task). The different experiments manipulated what participants saw on the screen as they were typing. Figure 1 summarizes the design of the three experiments. Experiment 1 had two conditions: the *word-feedback condition* was the baseline condition, in which participants saw what they typed on the screen in real-time, just like normal typing. In the *no-feedback* condition, nothing appeared on the screen until participant had finished typing and had responded yes/no in the metacognitive judgment task (Figure 1a). This experiment had two goals: 1) to replicate the finding of the critical dependence of corrections on visual feedback reported by Pinet & Nozari (2020) with a sample size more than twice as large. 2) To test the importance of visual feedback for conscious detection of errors (recall that this was not possible to determine in the previous study because we had only collected metacognitive judgments in the no-feedback condition).

To anticipate, results of Experiment 1 replicated a role for visual output in both conscious detection and correction of errors, albeit with a larger magnitude for the latter. This result brings up the question “What information in the visual output is so critical for applying repairs?”. The real-time visual output provides two types of information: *position* (where I am in the word?) and *letter identity* (what letter did I just type?). Removing positional information while retaining

letter identity is not feasible. Using a method such as rapid serial visual presentation (RSVP) of typed letters would remove spatial information about position but would retain temporal information. We thus opted for keeping positional information and removing letter identity using a condition similar to password typing.



*Figure 1: Schematics of the design of the three experiments. (a) In Experiment 1, a stimulus was presented auditorily, and participants typed it either with (word-f) or without (no-f) visual feedback. They made a metacognitive judgment on every trial and were shown what they had typed afterwards. The design of Experiments 2 (b) and 3 (c) differed from Experiment 1 only in the feedback provided during the typing period. Exp 2 included word-feedback and no-feedback conditions, and a new condition (position-f) condition in which letters were replaced with dots. Exp 3 included the word-f and position-f conditions and added a new (pos+cue-f) condition in which a red hash mark was displayed online on top of each erroneous letter typed. f = feedback; pos = position.*

Why would positional information help with correction? Correcting involves stopping the current action and deleting one or more letters to correct the erroneous keystroke. One role of visual information could be to provide information about the current position, in particular while going backwards during the correction process. If positional information is critical for corrections, independently of letter identity, we should see a significant improvement in repair

attempts when people have access to positional information in real-time as they type. Experiment 2 tested this prediction by adding the third condition, the *position-feedback condition*, in which dots appeared on the screen during typing, similar to when people type masked passwords in real life (Figure 1b). The results replicated the main findings of Experiment 1 and in addition revealed a small but statistically significant increase in correction rates when positional information was available in real time.

The results of Experiment 2 shed light on the relative contribution of positional information to corrections in real-life situations. But from a theoretical perspective, it can be argued that the absolute contribution of positional information to corrections can only be studied when error detection and error correction are cleanly disentangled. Experiment 3 was designed to address this issue. The *no-feedback* condition was removed, and a *position+cue-feedback condition* was added in which a red hash mark appeared immediately above the incorrect letter as soon as a participant typed it (Figure 1c). This condition, therefore, removed the need for *detecting* errors, by externally cueing participants to the occurrence of an error, and provided an opportunity to study the pure influence of positional information on corrections.

### **Application of the signal detection theory (SDT) to the studying of monitoring**

The current study follows the general theoretical framework proposed in Nozari & Hepner (2019a), which has been inspired by the signal detection theory (SDT; see Wixted, 2020; for a historical review and applications in cognitive psychology). The main assumption is that each attempt of producing a word, naturally leads to the generation of information that helps the system estimate the probability that the word might be an error. This information can be defined under the general umbrella of “conflict”, which is determined on the basis of the difference between the activation of two (or more) representations. When one representation has much stronger activation than all others, conflict is low, and the probability of an error is estimated to be low. When, on the other hand, more than one representation is activated, conflict is high, and the trial is estimated to have a higher chance of ending in an error. In this general framework, the hypothetical distribution of error trials is associated with an overall higher amount of conflict, while the hypothetical distribution of correct responses is associated with lower amounts of conflict (Figure 2). Since initiating a corrective action requires a binary decision (an error has or has not been committed), the system needs a way to translate the probability of error to a yes/no

decision. This is achieved by placing a criterion somewhere between the two distributions of correct and error trials. Trials associated with values of conflict higher than the criterion would be labeled as “error”, while trials with values of conflict lower than the criterion would be labeled as “correct” (Figure 2).

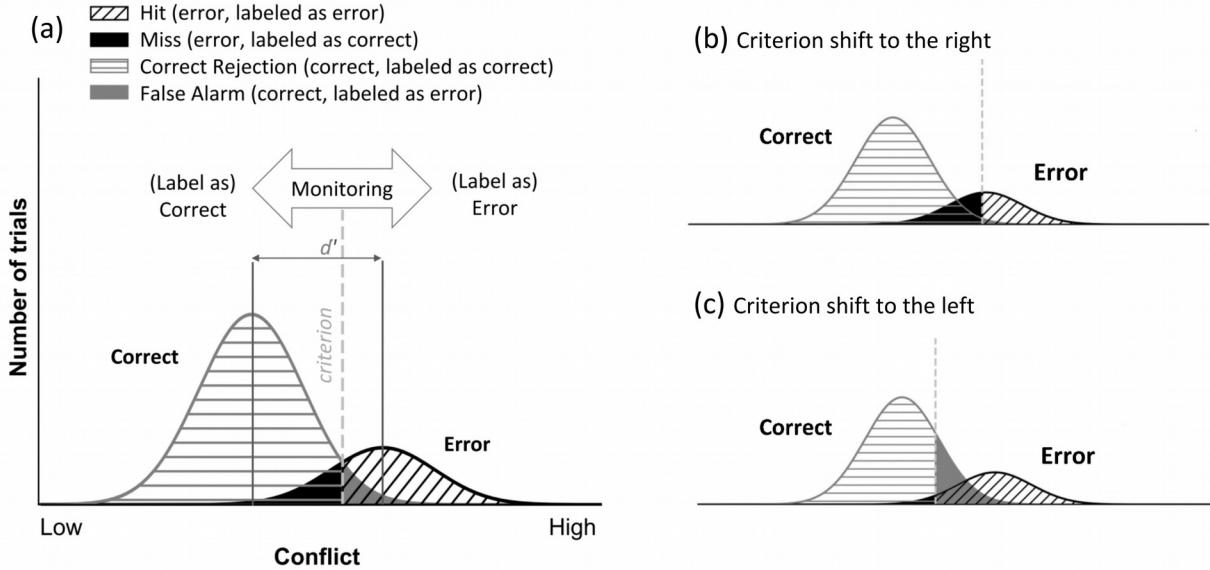
In earlier work (e.g., Hanley et al., 2016; Nozari et al., 2011; Nozari & Hepner, 2019a, 2019b), the source of conflict was primarily attributed to production-internal dynamics, i.e., activation of the correct and wrong response(s) within the production system. But conflict can also be defined in the context of other monitoring mechanisms. For example, in classic forward models (see Guenther, 2016 for a review), a motor target activates its corresponding sensory (auditory and proprioceptive) targets. When the actual auditory and proprioceptive stimulus (i.e., the outcome of production) is processed, if it is the same as the anticipated one, there is no conflict. If, however, the two differ, as would be in the case of errors, the system would experience “conflict”. Thus, the concept of conflict, broadly defined, can be used to unify different monitoring mechanisms under an umbrella of a multi-component monitoring system (Nozari, 2020). In this framework, the conflict signal is a sum of all sources of information provided by different monitoring mechanisms. This account makes a clear prediction: if a monitoring mechanism contributes significantly to monitoring, its removal should lead to a lower quality of the conflict signal, and consequently poorer monitoring performance. SDT provides a formal framework to test this hypothesis. Moreover, SDT can be applied separately to different tasks that are contingent on monitoring. We test two such behaviors, correction attempts and metacognitive judgments determining awareness over committed errors.

Four response types are defined in the SDT framework when applied to monitoring (also called SDT-type II; see Wixted, 2020 for a review): A Hit is defined as detecting/correcting a response which should have been detected/corrected, i.e., an error. A Miss is defined as not detecting/correcting what should have been detected and corrected, i.e., an error. A Correct Rejection refers to not detecting/correcting a response which did not need detection/correction, i.e., a correct response, and finally, a False Alarm refers to detecting/correcting a response which did not need detection/correction, i.e., a correct response. These four types of responses can be collected for both corrections and metacognitive judgments of conscious detection of errors. The SDT model fitted to individual participants’ data can estimate two critical parameters,  $d'$  (d prime), which indexes how far apart the two distributions are, and  $c$  (criterion), which indexes

where the criterion is placed for the binary (error/correct) decision (Figure 2a). Formally, this parameter denotes the number of standard deviations from the “ideal observer” position, i.e., when the criterion is placed at the intersection of the correct and error distributions. Therefore, positive values indicate a shift to the right, while negative values mark a shift to the left (Figure 2b and c).

When the quality of information is high, the two distributions have little overlap ( $d$  is large). A decrease in the quality of monitoring signal is thus expected to manifest as a lower  $d$ . In our experiments, we use this logic to measure the contribution of (full or partial) visual information to the monitoring signal. A significant drop in  $d$  would imply a substantial role for such information. The value of the criterion provides a different kind of information: how does behavior change as a function of decreased quality of information reflected in lower  $d$ ? Three outcomes are possible: 1) participants may shift their criterion to the right (Figure 2b). This reflects an attempt to decrease the probability of False Alarms, while sacrificing some Hits. 2) Alternatively, participants may shift their criterion to the left (Figure 2c). This means an attempt to keep the Hit rate high, while accepting an undesirable increase in False Alarms. 3) Finally, participants may keep their criterion in the same place, reflecting no particular preference in avoiding False Alarms or maximizing Hits.

In each of the three experiments, we estimated  $d$  and criterion by fitting a SDT model to data from individual participants separately for corrections and metacognitive judgments on error awareness, and used the results to draw conclusions about the contribution of visual information to the underlying monitoring signal in these two tasks, as well as changes to participants’ overt behavior as a function of removing such information.



*Figure 2.* Signal detection theory applied to the distributions of correct and error trials regarding the level of conflict. The distance between the two distribution is captured by  $d'$ . Larger  $d'$ s reflect cleaner separation of correct and error distributions, i.e., a cleaner monitoring signal. How to respond to the monitoring signal is determined by the placement of the criterion. The “ideal observer” position is assumed to be at the intersection of the two distributions (a). Criterion can be shifted either to the right (b) or to the left (c). Shifting the criterion to the right decreases False Alarms at the cost of increasing Misses. Shifting the criterion to the left has the opposite effect.

## Experiment 1

### Methods

#### Participants

In this and the following experiments, a predetermined sample size of 42 was set as the target. Participants were recruited via Amazon Mechanical Turk. They were consented under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine and were compensated for their participation. Participation eligibility was determined by a screening test administered before participants were accepted into the study. This test had two parts: In the first part, participants typed 15 words to dictation, with a relaxed deadline (5000 ms). In the second part, they typed 15 words to dictation, under a shorter deadline (2000 ms). The end of the auditory stimulus, i.e., when they were able to start typing was indicated by a short beep. The upcoming end of the typing period was indicated by a lower pitched beep, 500 ms before the effective end of the trial. Participants had to reach 80% accuracy on the first part, and finish

typing at least 80% of trials under the time limit (with minimum 50% accuracy) on the second part. If they did not meet these criteria, they did not proceed with the rest of the experiment. We ran as many participants as needed until 42 participants passed the screening test and were included. The total number of participants run to reach 42 who passed the test was 63.

The 42 participants (17 male) were 21-58 years old (mean =  $35.6 \pm 8$ ) and were all speaking English at home and at work (two were native Spanish speakers, that learned English at 3 and 9 years old)<sup>1</sup>. All participants had finished high school and 36 (86%) had some college education. Their typing speed and accuracy were estimated with a screening typing test at the beginning of the session (see below). In our sample typing speed ranged from 51 to 102 words per minute (mean =  $72.5 \pm 12$  wpm), and accuracy ranged from 80 to 100% (mean =  $93 \pm 6\%$ ).

### *Materials*

Stimuli were 600 7- and 8-letter words from the English Lexicon Project database (Balota et al., 2007). The number of items was determined by balancing the need for sufficient observations, while keeping the duration of the experiment under an hour. Word frequency (log) ranged from 1.7 to 3. Plural and compound words, and words that had homophones were not included in the material. Words were recorded by a native English speaker. They were divided into two lists of 300 items to be used in each condition, counterbalanced on number of phonemes, syllables, letters, phonological and orthographic neighbors, word and bigram frequency, and percentage of bimanual alternations.

### *Procedures*

The experiment was programmed using the jsPsych library (de Leeuw, 2015), embedded in an HTML environment. The Python library psiTurk (Gureckis et al., 2016) was used to handle participants' recruitment and compensation. Participants performed the experiment online using their own computer and keyboard. Experimental files are available at this address: 10.17605/OSF.IO/9TC5X. The main experimental session consisted of a typing-to-dictation task. We manipulated whether visual feedback appeared on the screen as participants were typing

---

<sup>1</sup>Since this study used a screening test to assess typing ability, the most unbiased decision was to include everyone who passed the screening test. To address potential concerns about including two non-native English speakers and one participant older than 40 years of age, we repeated all the analyses without these three participants. The pattern of results was unchanged. We, therefore, report the results using the whole sample.

(baseline/word-feedback condition) or not (no-feedback condition). In the no-feedback condition, the screen remained blank for the entire duration of the trial. Conditions were blocked and participants saw them in a counterbalanced order. One of the two balanced lists of stimuli was used for each feedback condition. The list used for each condition was also counterbalanced between participants.

The structure of a trial was as follows: participants heard a word, followed by a beep (1000 Hz, 100ms). They were instructed to type the word as fast and as accurately as possible when they heard the first beep and to finish typing before a second beep (500 Hz, 100ms), 1800ms after the first one. They had 500ms after the second beep to finish typing their answer, making it a total typing duration of 2300ms from the first beep before the trial was terminated. At the end of the trial, participants were prompted to make a metacognitive judgment on their performance, when the following sentence appeared: "Did you make an error in what you first typed?". They had to answer "yes" or "no" by pressing the "Y" or "N" keys, respectively. Participants were explicitly told that if they made a mistake and corrected it, it still counted as an error and they should answer "yes". Then there was a 500ms interval before the next word was presented. There were breaks every 50 trials.

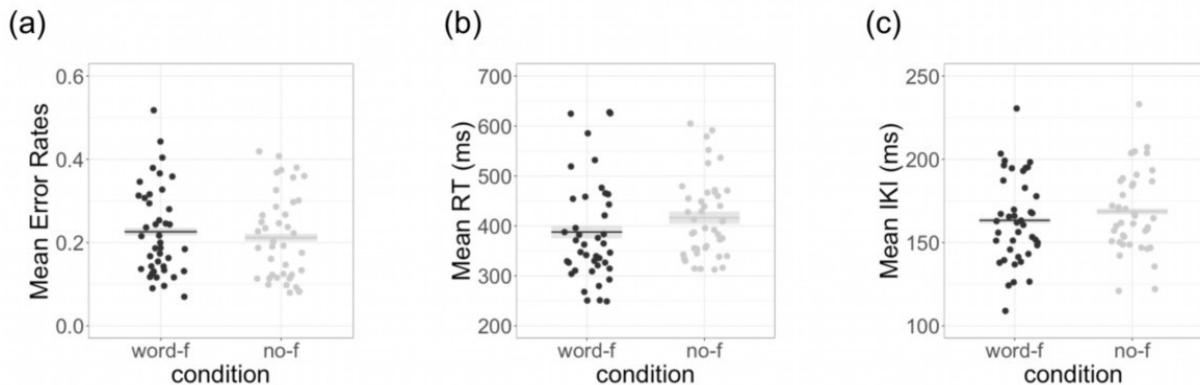
#### *Data analysis*

Any words that contained a keystroke not expected in the target word was considered an error. Reaction times (RTs) were calculated from the end of the auditory stimulus to the first keystroke. Interkeystroke intervals (IKIs) were the time intervals between consecutive keystrokes and were averaged over each word. RTs and IKIs outside of 3SD of the mean of each participant were removed for further analysis. Hit Rates, Misses, Correct Rejections and False Alarms were calculated for each condition for each participant, and SDT parameters ( $d'$  and criterion) were estimated using the *psycho* R package (version 0.4.91, Makowski, 2018, R version 3.3.3). General performance measures (error rates, RTs and IKIs) were analyzed using linear mixed effect models (*lmerTest* R package version 3.0-1, Kuznetsova et al., 2017). The main predictor was condition (word-feedback vs. no-feedback) using a simple effect coding. The random effect structure included random intercepts for both participants and items. As some models did not converge with the inclusion of random slopes, these were left out of all models to keep all analyses consistent. Because monitoring indices (SDT parameters) are calculated over all trials per participant, the structure of these data is no longer hierarchical. Therefore, these

parameters were analyzed using linear regression with similar contrast coding used for the analysis of general performance measures described above. Data and analysis scripts can be found here: 10.17605/OSF.IO/9TC5X.

## Results

One participant was excluded because his error rate was an outlier in the error distribution of participants based on visual inspection (above 65% in both conditions). From the remaining data, 2.4% of outlier RT and IKI data were excluded. Figure 3 shows the distribution of errors, RTs and IKIs in the word-feedback and no-feedback conditions. Across 41 participants, 5391 errors (22%) were committed, 2782 in the word-feedback and 2609 in the no-feedback condition. The error rate in the no-feedback condition was significantly lower than the word-feedback condition ( $21.2 \pm 10.3\%$  vs.  $22.6 \pm 10.7\%$ ;  $\beta = 0.088$ ,  $z = 2.6$ ,  $p = .009$ ). Average RTs, on the other hand, were higher in the no-feedback compared to the word-feedback condition ( $390.2 \pm 76\text{ms}$  vs.  $362 \pm 92\text{ms}$ ;  $\beta = 29.7$ ,  $t = 16.8$ ,  $p < .001$ ). The same was true for IKIs ( $168.7 \pm 24\text{ms}$  vs.  $163.3 \pm 26\text{ms}$ ;  $\beta = 5.86$ ,  $t = 15.2$ ,  $p < .001$ ).



*Figure 3. Mean error rates (a), RTs (b), and IKIs (c) in the word-feedback (word-f) and no-feedback (no-f) condition in Experiment 1.*

*Metacognitive judgments.* The overall rate of error awareness (Hit rate) was 69% and 54% in the word-feedback and no-feedback conditions, respectively (see Table 1). A comparison of Hit rates across the two groups showed significantly lower Hit rates in the no-feedback compared to the word-feedback condition ( $.53 \pm .13$  vs.  $.70 \pm .15$ ;  $\beta = -.18$ ,  $t = -5.8$ ,  $p < .001$ ; Figure 4a). Model-

estimated  $\bar{d}$  was significantly lower for the no-feedback compared to word-feedback condition ( $2.1 \pm 0.4$  vs.  $2.9 \pm 0.6$ ;  $\beta = -.81$ ,  $t = -7.0$ ,  $p < .001$ ; Figure 4b), whereas the location of the criterion did not significantly differ between the two conditions ( $1.0 \pm .4$  vs.  $.9 \pm .3$ ;  $\beta = .08$ ,  $t = 1.1$ ,  $p = .32$ ; Figure 4c).

*Corrections.* The overall rate of correction attempts (Hit rate) was 28% (763 attempts) in the word-feedback and 8% (221 attempts) in the no-feedback conditions (Figure 4c, Table 2). A comparison of Hit rates across the two groups showed significantly lower Hit rates in the no-feedback compared to the word-feedback condition ( $.09 \pm .10$  vs.  $.31 \pm .23$ ;  $\beta = -.22$ ,  $t = -5.8$ ,  $p < .001$ ; Figure 4e). Model-estimated  $\bar{d}$  was significantly lower for the no-feedback compared to word-feedback condition ( $1.2 \pm .6$  vs.  $2.0 \pm .8$ ;  $\beta = -.83$ ,  $t = -5.1$ ,  $p < .001$ ; Figure 4e). The difference in  $\bar{d}$ 's was accompanied by a criterion shift: the criterion was significantly higher in the no-feedback compared to the word-feedback condition ( $2.2 \pm .4$  vs.  $1.7 \pm .5$ ;  $\beta = 0.50$ ,  $t = 5.1$ ,  $p < .001$ ; Figure 4f).

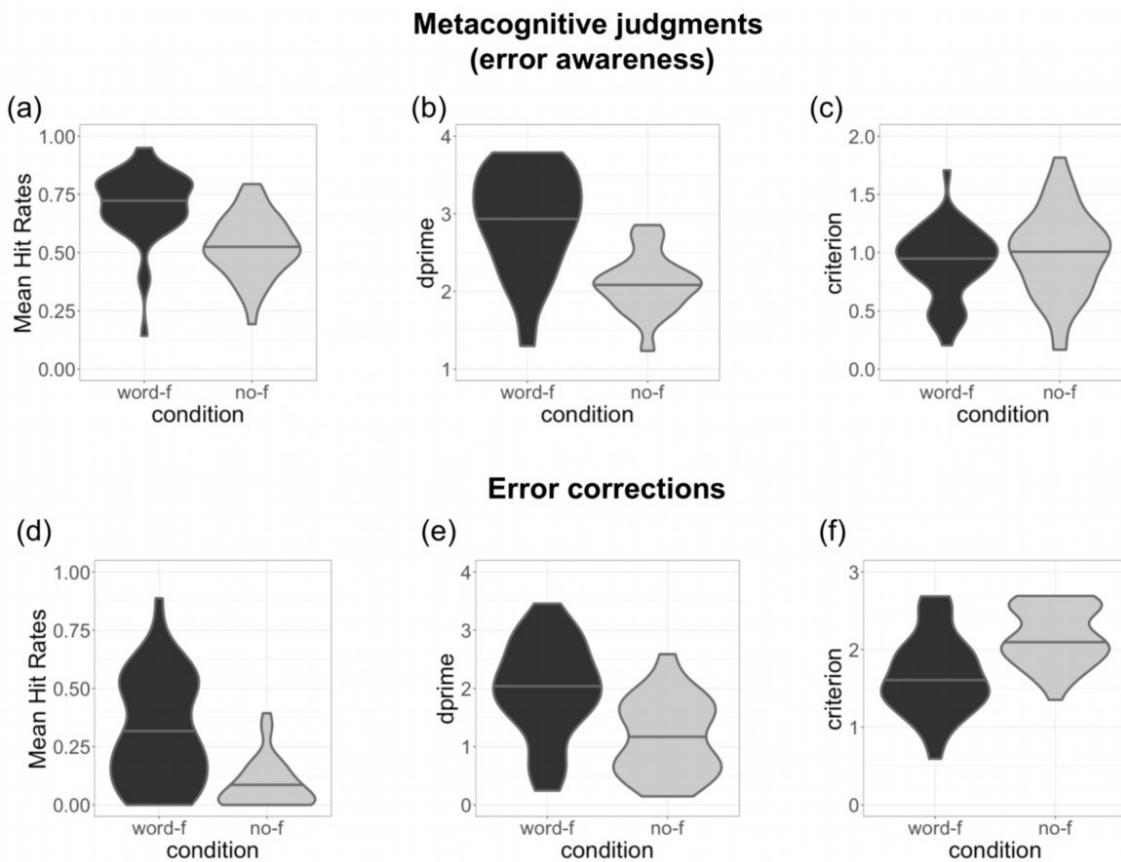


Figure 4. Results of the metacognitive judgments (upper panel) and corrections (lower panel) in the word-feedback (word-f) and no-feedback (no-f) conditions in Experiment 1.

Table 1. Descriptive statistics for error detection in the three experiments.

Experiment 1		Experiment 2			Experiment 3			
	Baseline	No	Baseline	No	Position	Baseline	No	Position
Raw counts	<i>Errors</i>	2782	2609	2050	1990	2041	1494	1406
	<i>Hit rate</i>	69%	54%	69%	55%	57%	72%	58%
By participants	<i>Hit rate</i>	$70.7 \pm 15\%$	$52.7 \pm 13\%$	$68 \pm 19\%$	$50 \pm 21\%$	$54 \pm 19\%$	$72 \pm 14\%$	$58 \pm 17\%$
	<i>FA rate</i>	$1.43 \pm 2.5\%$	$3.0 \pm 4.0\%$	$1.7 \pm 2.3\%$	$4.5 \pm 6.1\%$	$3.9 \pm 4.8\%$	$0.77 \pm 1.0\%$	$2.0 \pm 3.1\%$
	<i>Dprime</i>	$2.9 \pm 0.6$	$2.1 \pm 0.4$	$2.7 \pm 0.6$	$1.9 \pm 0.4$	$2.0 \pm 0.5$	$3.0 \pm 0.5$	$2.4 \pm 0.5$
	<i>Criterion</i>	$0.9 \pm 0.3$	$1.0 \pm 0.4$	$0.90 \pm 0.4$	$0.98 \pm 0.5$	$0.95 \pm 0.5$	$0.93 \pm 0.3$	$1.0 \pm 0.3$

*Table 2. Descriptive statistics for error correction in the three experiments.*

	Experiment 1		Experiment 2		Experiment 3			
	<i>Baseline</i>	<i>No</i>	<i>Baseline</i>	<i>No</i>	<i>Position</i>	<i>Baseline</i>	<i>No</i>	<i>Position</i>
Raw counts	<i>Errors</i>	763	221	403	94	195	558	259
	<i>Hit rate</i>	28%	8%	19%	3.90%	8.50%	35%	17%
By participants	<i>Hit rate</i>	$31 \pm 23\%$	$8.6 \pm 10\%$	$23 \pm 19\%$	$5.3 \pm 6\%$	$11 \pm 10\%$	$35 \pm 20\%$	$18 \pm 15\%$
	<i>FA rate</i>	$.35 \pm .6\%$	$.16 \pm .3\%$	$0.19 \pm 0.4\%$	$0.20 \pm 0.4\%$	$0.25 \pm 0.4\%$	$0.36 \pm 0.5\%$	$0.12 \pm 0.3\%$
	<i>Dprime</i>	$2.0 \pm 0.8$	$1.2 \pm 0.6$	$1.7 \pm 0.7$	$0.8 \pm 0.4$	$1.3 \pm 0.5$	$2.1 \pm 0.6$	$1.6 \pm 0.6$
	<i>Criterion</i>	$1.7 \pm 0.5$	$2.2 \pm 0.4$	$1.8 \pm 0.4$	$2.2 \pm 0.3$	$2.0 \pm 0.3$	$1.5 \pm 0.4$	$2.1 \pm 0.4$
								$1.9 \pm 0.5$

## Discussion

Removing immediate visual feedback caused a slowdown in initiation times and a lengthening of IKIs. These effects were accompanied by a slight but significant increase in accuracy. Thus, although these effects indicate a change in typing performance in the absence of immediate visual feedback, the presence of the speed-accuracy trade-off prevents the conclusion that efficient typing necessarily depends on the visual feedback. In fact, they suggest that accurate typing of single words is possible without visual feedback, although participants seem to approach the task with more vigilance, i.e., more stringent monitoring. Of specific interest to us was the role of visual feedback in error awareness and correction. Both suffered from the removal of immediate visual feedback, albeit to different degrees. While error awareness showed a 15% drop over a 69% baseline (word-feedback) condition (a 22% change), correction performance showed a 20% decrease over a 28% baseline condition (a 71% change). This suggests greater reliance of error correction —compared to error awareness— on visual feedback, replicating previous results from Pinet & Nozari (2020).

SDT fitting further shed light on the similarities and differences between conscious detection and correction: in both cases, there was a significant decrease in  $d'$  when visual feedback was removed. This result is compatible with the proposal that the conflict signal is a sum of multiple sources of information and gets noisier when one source (the external channel) is removed. On the other hand, a significant criterion shift was exclusive to correction attempts. This finding means that participants only change their correction (and not their judgment) behavior in response to the poorer discriminability in the no-feedback condition. This is most likely because correction attempts are costly; participants must stop ongoing behavior, erase the mistake, replace the segment with a repair, and resume typing. This takes both time and effort.

By shifting their criterion to the right (see Figure 2b), they are showing a tendency to minimize the chance of unnecessary corrections (False Alarms dropped from 3.5% to 1.6%). Yes/no metacognitive judgments, on the other hand, are not associated with similarly high cognitive costs. For this reason, participants need not shift their criterion despite the decreased discriminability.

These results raise an important question: why are corrections so difficult in the absence of visual feedback? In other words, what aspect of visual feedback provides the critical information necessary for correction attempts? Experiment 2 tests the contribution of two aspects of the visual signal to correction attempts: position in the sequence and letter identity.

## Experiment 2

### Methods

#### *Participants*

The recruitment, consenting, and compensation procedures were similar to Experiment 1. A total of 67 participants were screened until the pre-determined sample size of 42 was reached, none of whom had participated in Experiment 1. Participants (16 males) were 20-67 years old (mean =  $37.4 \pm 11$  years old) and all were native English speakers<sup>2</sup>. All participants had finished high school and 35 (85%) had some college education. The screening procedure was similar to Experiment 1. In the current sample typing speed ranged from 44 to 119 words per minute (mean =  $69.2 \pm 16$  wpm), and accuracy ranged from 80 to 100% (mean =  $93 \pm 6\%$ ).

#### *Materials*

Stimuli were similar to Experiment 1. The words were divided into three lists of 200 items each, counterbalanced for the number of phonemes, syllables and letters, word and bigram frequency, and percentage of bimanual alternations.

#### *Procedure*

The general procedure was similar to Experiment 1. In addition to the word-feedback and no-feedback conditions, we added a position-feedback condition. In this new condition, instead of the letters appearing on the screen, the symbol “•” appeared, similar to when typing a password online. Experimental files are available at this address: 10.17605/OSF.IO/9TC5X.

---

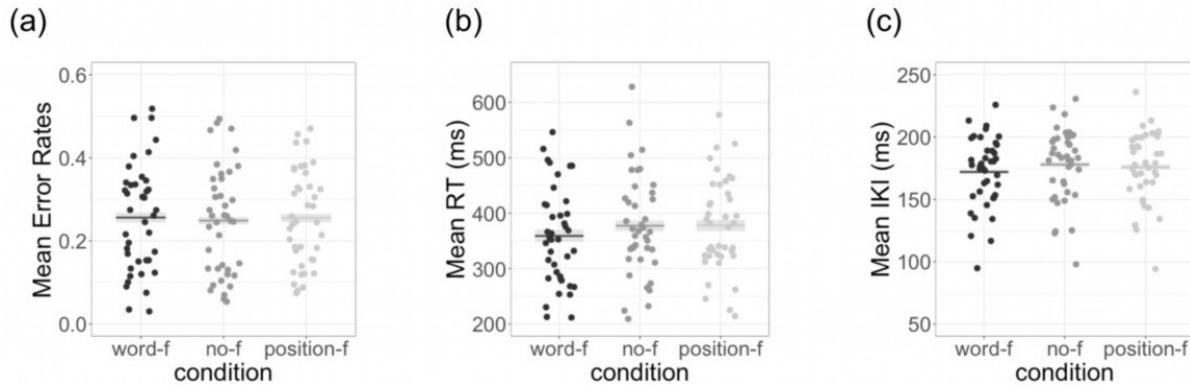
<sup>2</sup>Similar to Experiment 1, the removal of the oldest participant of the sample did not change the pattern of results reported here.

### *Data analysis*

We followed the same analysis procedure as Experiment 1. Contrasts were pre-planned and treatment-coded. To compare all our conditions to each other, we used two different contrasts codings, alternatively taking word-feedback and no-feedback as the reference level. Data and analysis scripts can be found here: 10.17605/OSF.IO/9TC5X

## **Results**

One participant was excluded because of a technical issue in recording its keystrokes. Another participant was removed because his mean RT was an outlier in the RT distribution of the sample based on visual inspection (above 900ms). Figure 5 shows the distribution of errors, RTs and IKIs in the word-feedback, no-feedback, and position-feedback conditions. Across 40 participants, 6081 errors (26%) were committed, 2050 in the word-feedback, 1990 in the no-feedback, and 2041 in the position-feedback condition. The error rate was not significantly different between the word-feedback and no-feedback conditions ( $25.6 \pm 13\%$  vs.  $24.9 \pm 13\%$ ;  $\beta = 0.035$ ,  $z = 0.86$ ,  $p = .39$ ), between the word-feedback and position-feedback ( $25.5 \pm 12\%$ ;  $\beta = -0.005$ ,  $z = -0.12$ ,  $p = .91$ ), or between the no-feedback and position-feedback ( $\beta = -0.042$ ,  $z = -0.98$ ,  $p = .33$ ). RTs, on the other hand, were significantly longer for the no-feedback compared to the word-feedback ( $377.3 \pm 87\text{ms}$  vs.  $359.2 \pm 87\text{ms}$ ;  $\beta = 19.6$ ,  $t = 8.8$ ,  $p < .001$ ). They were also significantly longer for the position-feedback compared to the word-feedback condition ( $378.0 \pm 84\text{ms}$ ;  $\beta = 21.2$ ,  $t = 9.6$ ,  $p < .001$ ). The comparison between the position-feedback and no-feedback revealed no significant difference ( $\beta = 1.6$ ,  $t = 0.72$ ,  $p = .48$ ). IKIs were also significantly longer for the no-feedback compared to the word-feedback ( $178.0 \pm 29\text{ms}$  vs.  $172.1 \pm 28\text{ms}$ ;  $\beta = 6.1$ ,  $t = 11.8$ ,  $p < .001$ ). The same was true for the comparison between position-feedback and word-feedback ( $176.0 \pm 28\text{ms}$ ;  $\beta = 4.2$ ,  $t = 8.3$ ,  $p < .001$ ). The comparison between position-feedback and no-feedback revealed significantly longer IKI in the no-feedback condition ( $\beta = -1.8$ ,  $t = -3.5$ ,  $p < .001$ ).

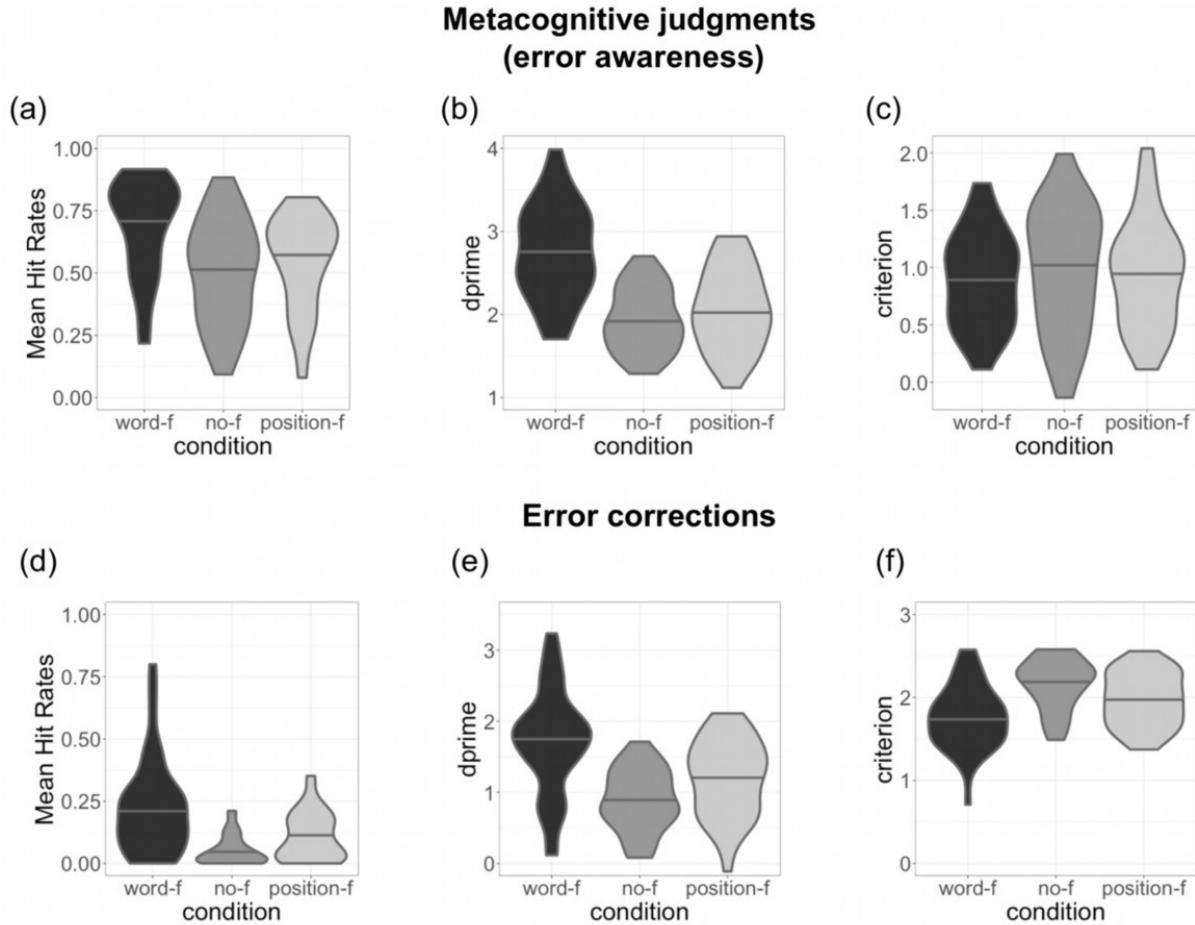


*Figure 5. Mean error rates (a), RTs (b), and IKIs (c) in the word-feedback (word-f), no-feedback (no-f) and position-feedback (position-f) condition in Experiment 2.*

*Metacognitive judgments.* The overall rate of error awareness (Hit rate) was 69% in the word-feedback, 55% in the no-feedback, and 57% in the position-feedback conditions, respectively (Table 1). Similar to Experiment 1, a comparison of Hit rates across the word-feedback and no-feedback conditions showed significantly lower Hit rates in the latter ( $.50 \pm .21$  vs.  $.68 \pm .19$ ;  $\beta = -0.18$ ,  $t = -4.0$ ,  $p < .001$ ; Figure 6a). The same was true in the comparison between the Hit rate in the position-feedback condition ( $.54 \pm .19$ ) vs. the word-feedback condition ( $\beta = -0.14$ ,  $t = -3.1$ ,  $p = .002$ ). There was no significant difference in the Hit rates between the position-feedback vs. no-feedback conditions ( $\beta = 0.039$ ,  $t = 0.88$ ,  $p = 0.38$ ). The contrasts comparing model-estimated  $d$ 's yielded similar results (Figure 6b). Average  $d$  was significantly lower in the no-feedback compared to word-feedback condition ( $1.9 \pm .4$  vs.  $2.7 \pm .6$ ;  $\beta = -0.81$ ,  $t = -7.5$ ,  $p < .001$ ). Similarly, average  $d$  in the position-feedback condition ( $2.0 \pm .5$ ) was significantly lower than word-feedback condition ( $\beta = -0.69$ ,  $t = -6.4$ ,  $p < .001$ ), but not significantly different from the no-feedback condition ( $\beta = 0.12$ ,  $t = 1.1$ ,  $p = .27$ ). The criterion, however, was not significantly different for any of the comparisons (word-feedback,  $.9 \pm .4$ , vs. no-feedback,  $.98 \pm .5$ ,  $\beta = 0.08$ ,  $t = .76$ ,  $p = .45$ ; word-feedback vs. position-feedback,  $.95 \pm .5$ ;  $\beta = 0.05$ ,  $t = 0.47$ ,  $p = .64$ ; no-feedback vs. position-feedback,  $\beta = -0.03$ ,  $t = 0.11$ ,  $p = .78$ ; Figure 6c).

*Corrections.* The overall rate of correction attempts (Hit rate) was 19% in the word-feedback, 4% in the no-feedback, and 9% in the position-feedback conditions, respectively (Table 2). Similar to Experiment 1, a comparison of Hit rates across the word-feedback and no-feedback

conditions showed significantly lower Hit rates in the latter ( $.23 \pm .19$  vs.  $.05 \pm .06$ ;  $\beta = -0.17$ ,  $t = -6.1$ ,  $p < .001$ ; Figure 6d). Hit rates in the position-feedback condition ( $.11 \pm .10$ ) were significantly lower than the word-feedback condition ( $\beta = -0.12$ ,  $t = -4.1$ ,  $p < .001$ ), and marginally higher than the Hit rates in the no-feedback condition ( $\beta = 0.06$ ,  $t = 2.0$ ,  $p = .0498$ ). Analyses of  $d'$  painted an even clearer picture: average  $d'$  (Figure 6e) was significantly lower in the no-feedback compared to word-feedback condition ( $.43 \pm .9$  vs.  $1.1 \pm 1.1$ ;  $\beta = -0.82$ ,  $t = -6.3$ ,  $p < .001$ ). Average  $d'$  in the position-feedback condition ( $.69 \pm .9$ ) was significantly different from both of the other conditions: significantly lower than the word-feedback condition ( $\beta = -0.51$ ,  $t = -3.9$ ,  $p < .001$ ), and significantly higher than the no-feedback condition ( $\beta = 0.31$ ,  $t = 2.4$ ,  $p = .019$ ). Moreover, in keeping with the results of Experiment 1, and in contrast to the pattern of results for metacognitive judgments, criterion placement changed as a function of condition (Figure 6f). The average criterion value was significantly higher in the no-feedback compared to word-feedback condition ( $2.2 \pm .3$  vs.  $1.8 \pm .4$ ;  $\beta = 0.39$ ,  $t = 5.1$ ,  $p < .001$ ). The average criterion value in the position-feedback ( $2.0 \pm .3$ ) was significantly higher than the word-feedback condition ( $\beta = 0.21$ ,  $t = 2.7$ ,  $p = .008$ ), and significantly lower than the no-feedback condition ( $\beta = -0.18$ ,  $t = -2.4$ ,  $p = .019$ ).



*Figure 6. Results of the metacognitive judgments (upper panel) and corrections (lower panel) in the word-feedback (word-f), no-feedback (no-f) and position-feedback (position-f) conditions in Experiment 2.*

## Discussion

Experiment 2 replicated most of the results reported in Experiment 1. RTs and IKIs were significantly longer in the no-feedback vs. word-feedback conditions. But unlike Experiment 1, accuracy was comparable between the two conditions, removing the concern about the speed-accuracy trade-off. The common finding of the two experiments, therefore, is that typing accuracy for single words in experienced typists does not critically depend on the presence or absence of visual feedback, but the removal of immediate visual feedback makes typists more cautious by slowing down the initiation and execution of the response (see also Pinet & Nozari, 2020). Importantly, the results of metacognitive judgments and corrections closely mirrored

those found in Experiment 1. The error awareness rate in the word-feedback and no-feedback conditions were very close in the two experiments (69% vs. 54% in Experiment 1 and 69% vs. 55% in Experiment 2), showing reliable and replicable estimates, with about 20% drop in error awareness as a function of the removal of immediate visual feedback in both cases. Error correction rates in the word-feedback and no-feedback conditions were noisier (28% vs. 8% in Experiment 1, and 19% vs. 4% in Experiment 2), but both showed large and comparable drops (71%-79%) as a result of the removal of immediate visual feedback. Critically, Experiment 2 replicated the much greater decline in the rate of error correction compared to error awareness in the absence of immediate visual feedback. Finally, Experiment 2 replicated the finding that participants' decision criterion only changed for corrections, but not for error awareness judgments in line with our interpretation that participants adjust their criterion based on the cost of a False Alarm, which is higher for repair attempts than for metacognitive judgments.

The new manipulation in Experiment 2 was providing immediate visual feedback in the forms of dots to mask the typed letters. This kind of feedback signals positional information to the participants without providing information about letter identity. Thus, changes to performance (if any) would reflect the role of positional information provided by visual feedback. The results were as follows: Not surprisingly, we found no evidence to suggest that providing positional information helped error awareness; performance in the metacognitive judgment task was comparable between no-feedback and position-feedback conditions, and both were significantly poorer than the word-feedback condition with comparable decline rates (20% for no-feedback and 17% for position-feedback). We can thus conclude that becoming aware of having committed an error shows some sensitivity to the presence of immediate visual feedback, but the critical information is not about the position, but the identity, of letters.

In contrast to this, we found that providing positional information helped correction performance; repair Hit rates were doubled in the position-feedback compared to no-feedback condition (9% compared to 4%), although they did not quite reach the rate of repair in the word-feedback condition (19%). Analyses of  $d$ 's showed that the signal provided by position in the position-feedback condition was significantly better than the no-feedback, but significantly poorer than the word-feedback condition. Analyses of the criterion values mirrored these findings; the criterion in the position-feedback condition falls right in between the no-feedback and word-feedback conditions, meaning that participants were encouraged by the positional

information to attempt more repairs, but were still aware of the higher chance of False Alarms compared to when both position and identity information were available, thus not shifting their criterion all the way to the level matching the word-feedback condition.

To summarize, Experiment 2 replicated the key findings in Experiment 1, and shed more light on the way in which visual information may be important for typing (Table 3). The results indicated that error correction (but not error awareness) benefited from the presence of positional information, but such information alone was not sufficient to promote correct repairs to the level of full visual feedback with letters identity. The application of SDT further showed that participants were keenly aware both of the usefulness of positional information and of the limits provided by such information, and adjusted their criterion accordingly.

While these results show the limitation of positional information in guiding corrections in settings that are applicable to real-life (e.g., password typing), we cannot claim that we have determined the absolute limits of the contribution of visual information to correction performance. To answer that question, we must create a situation in which positional information has the best possible chance of helping typing performance. Experiment 3 does so by providing deterministic external cues when an error is made (i.e., removing uncertainty about “whether” a correction is necessary or not), and comparing correction performance with positional information alone vs. normal feedback, which contains information on both the position and identity of typed letters. If participants’ corrections improve with external cueing, we would conclude that the lower correction rates were due to uncertainty about whether there is an error. If not, we would conclude that the fewer correction attempts reflect participants’ estimation that there is not sufficient information to undertake a repair, regardless of detecting one.

## Experiment 3

### Methods

#### *Participants*

The recruitment, consenting, and compensation procedures were similar to Experiments 1 and 2. A total of 80 participants were screened with the aim of including 42 participants who passed the screening test. Due to a registration error, only 41 participants were included in this study, none of whom had participated in either Experiments 1 nor 2. Participants (17 male) were 24-55 years old (mean =  $35.7 \pm 7$ ) and all were native English speakers. All participants had

finished high school and 35 (82%) had some college education. The screening procedure was similar to Experiments 1 and 2. In the current sample typing speed ranged from 46 to 101 words per minute (mean =  $75.7 \pm 14$  wpm), and accuracy ranged from 80 to 100% (mean =  $94 \pm 6\%$ ).

### *Materials*

Stimuli were similar to Experiment 1 and 2. The same three lists of 200 items as in Experiment 2 were used.

### *Procedure*

The general procedure was similar to Experiment 1. We removed the no-feedback condition, and instead added a *position+cue-feedback* condition to the word-feedback and the position-feedback conditions. The new position+cue-feedback condition was similar to the position-feedback condition, in that the symbol “•” appeared instead of the letters when participants were typing, but had the additional feature of displaying a red hash mark above the dot whenever participants typed the wrong letter. Experimental files are available at this address: 10.17605/OSF.IO/9TC5X.

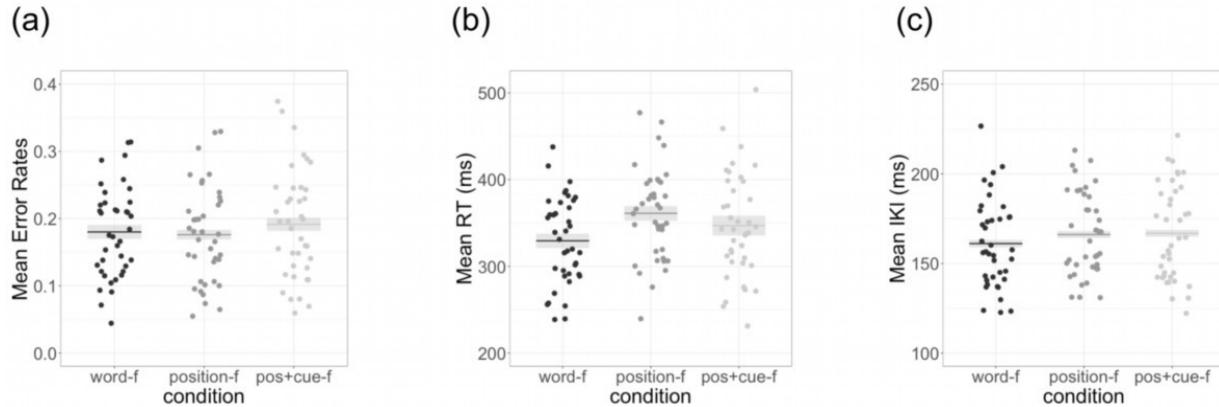
### *Data analysis*

We followed the same analysis procedures as in Experiment 2. Contrasts were pre-planned and treatment-coded. To compare all our conditions to each other, we used two different contrasts settings, alternatively taking word-feedback and position-feedback as the reference level. Data and analysis scripts can be found here: 10.17605/OSF.IO/9TC5X.

## **Results**

Two participants were excluded, one because of his outlier error rate, and one because of his outlier average RT, based on visual inspection of the data. Figure 7 shows the distribution of errors, RTs and IKIs in the word-feedback, position-feedback, and position+cue-feedback conditions. Across 39 participants, 4274 errors (18%) were committed, 1494 in the word-feedback, 1406 in the no-feedback, and 1216 in the position-feedback condition. The error rate was not significantly different between the word-feedback and position-feedback conditions ( $19.2 \pm 8\%$  vs.  $18.0 \pm 7\%$ ;  $\beta = 0.068$ ,  $z = 1.5$ ,  $p = .13$ ), but it was significantly lower in the position+cue-feedback ( $17.6 \pm 7\%$ ) compared to the word-feedback position ( $\beta = 0.11$ ,  $z = 2.5$ ,  $p = .01$ ). Position-feedback and position+cue-feedback conditions did not differ significantly in

terms of accuracy ( $\beta = 0.05$ ,  $z = 1.2$ ,  $p = .21$ ). Responses were significantly slower for the position-feedback compared to the word-feedback ( $361.1 \pm 52$  vs.  $329.4 \pm 50$ ;  $\beta = 32.8$ ,  $z = 15.9$ ,  $p <.001$ ) condition. They were also significantly slower for the position+cue-feedback ( $347.0 \pm 60$ ) compared to the word-feedback condition ( $\beta = 18.9$ ,  $z = 9.2$ ,  $p <.001$ ). The position+cue-feedback was, however, significantly faster than the position-feedback ( $\beta = -13.9$ ,  $z = -6.7$ ,  $p <.001$ ). IKIs were also significantly longer for the position-feedback compared to the word-feedback ( $166.2 \pm 23$ ms vs.  $161.1 \pm 25$ ms;  $\beta = 5.4$ ,  $z = 11.0$ ,  $p <.001$ ). The same was true for the comparison between position+cue-feedback ( $166.8 \pm 26$ ms) and word-feedback ( $\beta = 6.0$ ,  $z = 12.2$ ,  $p <.001$ ). The comparison between position+cue-feedback and position-feedback revealed no significant difference ( $\beta = 0.57$ ,  $z = 1.16$ ,  $p = .25$ ).



*Figure 7. Mean error rates (a), RTs (b), and IKIs (c) in the word-feedback (word-f), position-feedback (position-f), and position+cue-feedback (pos+cue-f) condition in Experiment 3.*

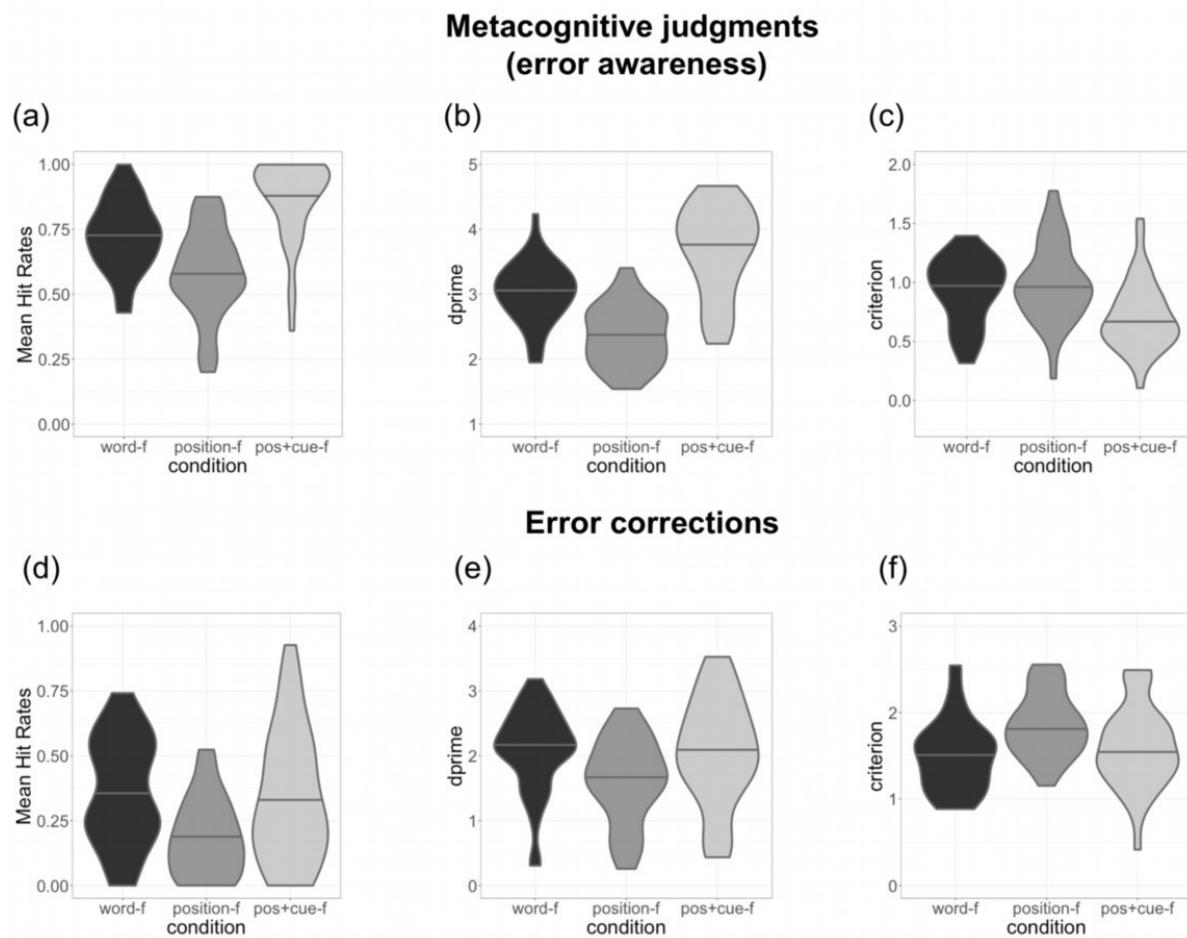
*Metacognitive judgments.* The overall rate of error awareness (Hit rate) was 72% in the word-feedback, 58% in the position-feedback, and 86% in the position+cue-feedback conditions, respectively (Table 1). Similar to Experiment 2, a comparison of Hit rates across the word-feedback and position-feedback conditions showed significantly lower Hit rates in the latter ( $.72 \pm .14$  vs.  $.58 \pm .17$ ;  $\beta = -0.14$ ,  $t = -4.0$ ,  $p <.001$ ; Figure 8a). Not surprisingly, error awareness was the highest in the position+cue-feedback condition, since the presence of an error had been externally cued. Performance was still not perfect ( $.87 \pm .16$ ), indicating that participants must have missed some of the error cues. Nevertheless, metacognitive judgments of error awareness were significantly higher in this condition compared to both the word-feedback ( $\beta = 0.14$ ,  $t = 4.1$ ,

$p < .001$ ) and position-feedback ( $\beta = 0.28$ ,  $t = 8.1$ ,  $p < .001$ ) conditions. Since the cueing of errors has been externally provided in the position+cue-feedback condition, it is not meaningful to compare  $d$  and criterion for metacognitive judgments in this condition compared to the other two. However, the comparison of  $d$ 's and criterion placement between word-feedback and position-feedback conditions is meaningful, and replicates what was reported in Experiment 2. Average  $d$  was significantly lower in the position-feedback ( $2.4 \pm .5$ ) compared to the word-feedback condition ( $3.0 \pm .5$ ;  $\beta = -0.66$ ,  $t = -5.2$ ,  $p < .001$ ; Figure 8b). The criterion was, on the other hand, not significantly different between the two conditions (word-feedback,  $.93 \pm .3$ ; vs. position-feedback,  $1.0 \pm .3$ ;  $\beta = 0.07$ ,  $t = 0.96$ ,  $p = .34$ ; Figure 8c).

*Corrections.* The overall rate of correction attempts (Hit rate) was 35% in the word-feedback, 17% in the position-feedback, and 33% in the position+cue-feedback conditions, respectively (Table 2). Hit rates in the position-feedback condition ( $.19 \pm .15$ ) were significantly lower than the word-feedback condition ( $.36 \pm .21$ ;  $\beta = -0.17$ ,  $t = -3.7$ ,  $p < .001$ ), whereas Hit rates in the position+cue-feedback condition ( $.34 \pm .25$ ) were comparable to the word-feedback condition ( $\beta = -0.02$ ,  $t = -0.5$ ,  $p = .62$ ), and significantly higher than the position-feedback condition ( $\beta = 0.15$ ,  $t = 3.3$ ,  $p = .0013$ ). Analysis of  $d$ 's found an identical pattern (Figure 8e), with significantly lower  $d$ 's in the position-feedback compared to the word-feedback condition ( $1.6 \pm .6$  vs.  $2.1 \pm 0.6$ ;  $\beta = -0.47$ ,  $t = -2.9$ ,  $p = .004$ ), comparable  $d$  rates in the position+cue-feedback ( $2.1 \pm 0.8$ ) and word-feedback conditions ( $\beta = -0.02$ ,  $t = -0.15$ ,  $p = .88$ ), and significantly higher  $d$ 's in the position+cue-feedback compared to the position-feedback condition ( $\beta = 0.44$ ,  $t = 2.8$ ,  $p = .0059$ ). Criterion was significantly higher in the position-feedback ( $2.1 \pm .4$ ) compared to the word-feedback condition ( $1.5 \pm .4$ ;  $\beta = 0.35$ ,  $t = 3.7$ ,  $p < .001$ ; Figure 8f), whereas criterion in the position+cue-feedback condition ( $1.9 \pm .5$ ) was comparable to the word-feedback condition ( $\beta = 0.09$ ,  $t = 0.9$ ,  $p = .35$ ), and significantly lower than the position-feedback condition ( $\beta = -0.26$ ,  $t = -2.9$ ,  $p = .0046$ ).

*Successful repairs.* The results of the correction analysis showed that correction attempts were bumped up to the level of word-feedback condition when the occurrence of an error was externally cued. This might simply reflect the pressure to correct because of the external cue. We thus conducted an additional analysis to understand whether these externally cued repairs had the

same success rate in correcting the mistake as the self-initiated repairs. Of the committed corrections, 295 (53%) were successful repairs in the word-feedback, 142 (55%) in the position-feedback, and 180 (41%) in the position+cue-feedback condition. A non-parametric test showed that, despite comparable Hit rates for correction attempts, the success rate was significantly lower for the externally-cued repairs (i.e., in the position+cue-feedback) compared to the word condition ( $40\pm21\%$  vs.  $50\pm24\%$ ;  $z = -2.58$ ,  $p = .0096$ ). On the other hand, while position-feedback had a lower Hit rate than the word-feedback condition, the success rate in the former ( $47\pm28\%$ ) was comparable to the latter ( $z = -0.39$ ,  $p = .69$ ).



*Figure 8. Results of the metacognitive judgments (upper panel) and corrections (lower panel) in the word-feedback (word-f), position-feedback (pos-f) and position+cue-feedback (pos+cue-f) conditions in Experiment 3.*

## Discussion

Comparisons of error rates and RTs between word-feedback and position-feedback replicated those of Experiment 2. Error rates in the position+cue-feedback condition was comparable to that in the position-feedback condition, but participants were faster in the former, most likely because the external error signals reduced their reliance on internal monitoring processes that slowed them down in the visually-impoveryed typing condition. Comparisons of metacognitive judgments of error awareness in the word-feedback and position-feedback conditions also replicated those reported in Experiment 2, with a modest but significant drop in error awareness when letter identity has been removed. We can thus conclude with confidence that visual information in the form of letter identity is important for error awareness. As for corrections, in both experiments Hit rates were significantly lower in the position-feedback compared to the word-feedback condition. Analyses of the SDT-derived parameters confirmed this pattern. In both experiments,  $d$  was significantly lower for the position-feedback condition and the criterion was significantly higher, implying the same tendency to attempt fewer repairs in the condition with lower discriminability, i.e., the position-feedback condition, in both experiments. Together, these findings suggest that positional information does not help error awareness, but does improve corrections, albeit not to the level that full visual information does. Moreover, when the success of repair attempts was compared, we found comparable success rates between word-feedback and position-feedback conditions (48% vs. 46%), showing that although the initiation of the repair depended, to some extent, on seeing the letters, once a repair was internally initiated, it was completed with equal success rates independently of visual information regarding the letter identity.

Recall that error awareness was lower in the position-feedback compared to the word-feedback condition (Table 3). If this is the source of fewer correction attempts in the position-feedback conditions, externally cueing errors should significantly improve the correction behavior. The position+cue-feedback condition in Experiment 3 was designed to test this prediction. At first glance, it seems that the results confirmed this prediction: correction performance in the position+cue-feedback condition rose to the level of word-feedback condition. But testing the success of repairs told a different story: when errors were externally cued, only 40% of the attempted repairs were successful, which was significantly lower than the

success rate in the word-feedback condition. To summarize, fewer correction attempts in the absence of visual information are not primarily due to the uncertainty about whether an error has or has not occurred. Rather, the lower correction attempts in the absence of visual information about letter identity imply that participants rely on this information and can gauge that repairs are much less likely to be successful in the absence of such information.

*Table 3. Qualitative summary table of the findings in the three experiments. Higher/lower stand for the direction of significant differences, N.S. stands for non-significant difference. IKI = Interkeystroke interval, RT = Response time; Pos = position.*

		Experiment 1	Experiment 2		Experiment 3		
		No vs. No vs. Baseline	Position vs. Baseline	Position vs. No	Position vs. Baseline	Pos+Cue vs. Baseline	Pos+Cue vs. Position
Performance	Error rate	lower	N.S.	N.S.	N.S.	lower	N.S.
	RT	higher	higher	higher	higher	higher	lower
	IKI	higher	higher	higher	higher	higher	N.S
Error detection	Hit rate	lower	lower	lower	lower	higher	higher
	Dprime	lower	lower	lower	N.S.	lower	-
	Criterion	N.S.	N.S.	N.S.	N.S.	-	-
Error correction	Hit rate	lower	lower	lower	higher	lower	N.S.
	Dprime	lower	lower	lower	higher	lower	N.S.
	Criterion	higher	higher	higher	lower	higher	lower

## General Discussion

This study investigated the role of visual information provided by the external channel on monitoring performance in typing. The typing-to-dictation paradigm taps into lexical-to-

segmental mapping and is thus a good choice for studying segmental errors which result from this part of the language production system. Using this task, in three experiments, we manipulated visual feedback by removing part of or all visual information from the screen, and investigated the change to typing speed and accuracy, as well as conscious detection and correction of errors. In a nutshell, results suggested that removal of the visual feedback changed monitoring performance by slowing down the initiation and execution of typing, causing a modest decrease in conscious awareness of errors, and a much more dramatic decrease in the correction of errors. While providing positional information (as in masked password typing) caused a statistically significant improvement in corrections, the effect was small, suggesting that participants relied heavily on letter identity for applying repairs. We unpack these findings below.

### **The effect of visual feedback on the speed and accuracy of typing**

When visual feedback was completely removed in Experiment 1, participants were slower to start typing (i.e., longer RTs) and took longer to type the whole word (longer IKIs). Longer RTs and/or IKIs as a function of removing visual feedback replicate previous findings in both typing to dictation (Pinet & Nozari, 2020) and copy-typing (Snyder et al., 2015), and were internally replicated in this paper as well. Reports on the change to typing accuracy, however, have been less consistent. Snyder and Logan (2015) reported a significant decrease in the accuracy of copy-typing when visual feedback was removed, but that task is likely to exaggerate reliance on visual information. Of the three experiments that have used typing to dictation with similar methodologies, two have found no change in accuracy (Pinet & Nozari, 2020, and Experiment 2 in the current paper), while one found statistically significantly higher accuracy in the absence of visual feedback (Experiment 1 in the current paper). These results suggest that typing accuracy in practiced typists does not critically depend on the visual outcome, at least for single words. However, the absence of such information consistently changes the speed with which people approach and complete the task (i.e., both reaction times and typing speed). This is compatible with a multi-channel monitoring account in which the removal of part of the information pertinent to monitoring increases the uncertainty about performance outcome. Slowing down is a successful coping strategy, since it allows for greater accumulation of evidence via the internal monitoring channels (see Nozari & Hepner, 2019a for a discussion).

### **The effect of visual feedback on conscious detection and correction of errors in typing**

While prior studies had investigated the effect of removing visual feedback on typing, two issues had remained unaddressed: (a) the first concerns the separation of the effect on conscious detection of errors vs. error correction. This distinction is especially important in light of the evidence pointing to some degree of automaticity in error correction, such as quick and efficient repairs in young children and individuals with aphasia that is dissociable from conscious awareness of the nature of the error (see Nozari, Martin, & McCloskey, 2019 for a discussion of evidence). Prior studies either did not probe conscious error detection (Kalfaoglu & Stafford, 2014), did not have the correct condition for distinguishing error awareness from correction (Pinet & Nozari, 2020), or prohibited corrections altogether (Snyder et al., 2015). (b) The second issue was that while all these studies hinted at some influence of visual information on monitoring performance, it remained unclear what kind of visual information was critical for good monitoring behavior. The present study addressed these two concerns.

Results of both Experiments 1 and 2 showed that correction rates plummeted when the visual word was completely removed, with a dramatic decrease in the range of 70-80%, closely replicating the results of Pinet & Nozari (2020). Previously, Kalfaoglu & Stafford (2014) reported higher correction rates (~60%) in the absence of visual feedback, but it must be noted that their study provided no baseline with normal visual feedback against which this number can be compared. Importantly, participants were explicitly instructed to correct as many errors as they could with no time limit on the task, and copied the sentences they had to type, meaning that they had access to correct spelling for all the words. Given that these conditions should provide the best chance for corrections, a correction rate of only 60% implies a critical role for visual feedback in monitoring, compatible with the conclusion drawn from the findings of the experiments reported in the current paper.

In comparison to correction attempts, conscious detection of errors was less sensitive to the removal of visual information. Still detection rate did decrease by about 20% in both Experiments 1 and 2. In line with our interpretation that removing the external channel deprives the system of one source of information useful for distinguishing between correct and error trials, application of SDT to the data revealed significantly smaller  $\delta$  (i.e., less cleanly-separated correct and error distributions) in the absence of visual information for both correction and conscious

detection of errors. Interestingly, the decrease in  $\bar{d}$ , common to both correction and conscious detection, was accompanied by different patterns of criterion setting. Experiment 1 showed that participants only adjusted their criterion for corrections (a shift towards minimizing False Alarms) but did not change its location for metacognitive judgments. Experiment 2 replicated this pattern.

The dissociation between the pattern of change to  $\bar{d}$  and  $c$  for corrections vs. conscious detections demonstrates that while the underlying mechanisms for both functions make use of visual information, and thus suffer from the removal of such information, the system responds differently to the decreased quality of information in these two cases. When False Alarms are costly, e.g., as in attempting a repair that was not needed, participants are much more willing to sacrifice a Hit when the quality of information for separating correct and error trials is poor. On the other hand, when there is no increase to the cost, as in no consequence to responding yes/no in a metacognitive judgment task, participants are willing to accept the increased False Alarm rate in order to avoid a large increase in Misses. This pattern, in turn, suggests that the great decrease in correction rates when visual information is absent, is less about not knowing that an error has been committed (and thus a repair was needed), and more about avoiding the cost of an uncertain repair.

The above pattern naturally leads to the second issue that has been overlooked in past studies, i.e., what kind of information do participants rely on for corrections, the absence of which causes such large drops in correction rates? The visual word contains two potentially useful sources of information as feedback to the typist: position and letter identity. Experiment 2 removed letter identity while preserving positional information by creating a masked-password-like typing situation. The first finding was that providing positional information did not help conscious awareness of errors and there was no reason to expect it would. This finding was replicated in Experiment 3. Providing positional information, however, caused a small but statistically significant increase in correction rates. Average  $\bar{d}$  increased compared to the no-feedback condition, suggesting that positional information provided a useful source of information for attempting correction. Criterion also shifted to the left compared to the no-feedback condition, suggesting that with the addition of more information participants were more willing to attempt correction in order to increase the Hit rate. Both  $\bar{d}$  and criterion fell between the word-feedback and no-feedback conditions, suggesting indirectly that while positional

information was helpful, letter identity remained an important source of information for corrections. Finally, Experiment 3 was designed to determine whether the drop in correction rates in the absence of letter identity reflected participants' hesitation about whether an error had occurred, or alternatively, participants' accurate evaluation that the current information was not sufficient to attempt a successful repair. By externally cueing errors, the task removed the uncertainty about error occurrence. The results showed that this manipulation prompted more repair attempts, but critically, these repairs were less successful than those that were self-initiated in conditions without external error cues. This finding shows that the fewer repair attempts in the absence of letter identity is a sound decision reflecting the lack of sufficient information for repairs that have a high likelihood of success. This, in turn, suggests that visual information about letter identity plays a key role in successful corrections.

A potential concern in this study is that removing visual feedback may have shifted attention towards visual inspection of fingers during typing, and thus visual information may have continued to contribute significantly to the detection and repair processes. There are several reasons to believe that the results are not strongly dependent on this possibility. (1) The average typing speed in this sample was 72 wpm, which leaves little time for detailed processing of visual information from fingers, even if participants have looked at their hands. (2) In self-reports administered at the end of each experiment fewer than 1/3 of participants reported looking at their hands during typing (26% for Experiment 1, 29% for Experiment 2 and 3). Note that there was no penalty for reporting that they did look at their hands during typing. Therefore, there is no reason to believe that the self-reports are unreliable. (3) Importantly, the results suggest that the removal of visual feedback from the screen *did* have a statistically significant effect on both detection and correction. If these effects were observed despite the fact that participants may have compensated for some of the lost visual information by looking at their hands, then the current results constitute even stronger evidence for the role of visual feedback from the screen on error awareness and repair behavior. It is also worth mentioning that covering hands to avoid the possibility of looking at them can cause other issues, such as misplacing the hands on the keyboard, which would lead to errors that are not cognitively interesting. Therefore, even though we cannot rule out the possibility that some of the participants may have looked at their hands on some of the trials, the current design is still the most suitable for our purpose, and the conclusions are unlikely to change based on this possibility.

We also acknowledge that the findings of the study were obtained from a population of typists who demonstrated a certain level of typing proficiency (as tested by our screening tests). It is possible (and perhaps expected) that visual feedback plays a much more critical role in the speed and accuracy of typing performance, and potentially also on monitoring performance, in less proficient typists.

### **Implications for theories of monitoring and control in language production**

Traditionally, language research has remained somewhat separate from research on the rest of cognition, due in part to the influence of modular views (e.g., Fodor, 1983) and in part, the linguistic theory (e.g., Chomsky, 1968). Consequently, a fundamental problem in cognitive science, i.e., the problem of task optimization and the frameworks that have been developed to address this problem, have made little contact with the field of language production (Anders & Alario, 2019; Anders, Riès, van Maanen, & Alario, 2015; see Nozari & Hepner, 2019a for a criticism). This omission has influenced theoretical positions on some of the fundamental aspects of language production. For example, Nozari and Hepner (2019b) demonstrated that the long-standing question around the competitiveness of lexical selection is ill-posed if the problem of selection is viewed as a SDT-type II problem. This, in turn, changes the question from “is lexical selection competitive or not?” to “what factors determine the degree of competitiveness in the lexical selection process?”, and provides a path by investigating the distinct contribution of the factors that affect the underlying information vs. those that affect the placement of a selection criterion (see also Costa, 2019; Mahon & Navarrete, 2019; Melinger & Abdel Rahman, 2019; Nozari & Hepner, 2019a; Oppenheim & Balatsou, 2019).

A very similar problem exists in how language research has viewed monitoring and control mechanisms. For decades the dominant monitoring account was a comprehension-based account (Levelt, 1983; 1989) which, by the virtue of being based on “comprehension”, viewed the conscious detection and correction of errors as the goal of the monitoring process. Aside from the problem of timing mentioned in the introduction, this account does not naturally predict a dissociation between metacognitive awareness over errors and a correction attempt, i.e., if the goal of the monitoring system is to correct errors, why doesn’t it do that job, and what is the advantage of being aware of an error if it is not corrected? Approached from the perspective of a task-optimization framework, however, the problem makes much more sense: what contributes

to decisions (implicit or explicit) regarding conscious detection and correction is determined a) by sources of information that determine the shape of the underlying distributions of correct and error responses, and b) by the variables that determine the location of the decision criterion.

The implication of adopting this perspective is posing a new set of questions as key questions for understanding the monitoring-control loops that regulate language production. For example, expanding the scope of monitoring from a system that detects occasional “errors” to one that constantly monitors the subtle fluctuations in control demand in the system in order to adjust the level of control (Freund & Nozari, 2018), shifts the focus from questions such as which error detection mechanism is the *right one* (Roelofs, 2020), to investigating the contribution of various sources of information that are relevant for detecting subtle changes in system’s performance. The current study provided a framework for doing just that. Moreover, the separation of detection/correction functions from metacognitive awareness over errors paves the way for investigating repair mechanisms that have not been entertained in the language production literature until recently. For instance, Logan (2018) asserts that despite automaticity in the primary production processes, repairs must depend on attentional control. Contrary to this view, Nozari et al. (2019) and Nooteboom & Quené (2020) propose that the basic repair mechanism need not be dependent on conscious and controlled processes. Taking advantage of the fact that multiple related responses are often activated during production, these authors proposed that correct responses are likely to be highly co-activated along with the error responses. If an error is detected, e.g., through having activation levels close to a competing response, the system may automatically select the next most highly activated response as repair. In support of this idea, Nooteboom & Quené (2020) showed that repairs were often the correct response and more so in cases where competition from other responses was minimal, as in the case of single errors. These errors are also shown to have quicker error-to-cutoff and cut-off-to-repair times, further showing the quick process of swapping them with the next best response, compared to messier errors that have more than one competitor activated. The dissociation between the correction behavior and metacognitive awareness observed in the current paper suggests that conscious awareness of having committed an error is certainly not enough for attempting repairs and may not even be necessary for initiating a repair. This, in turn, prompts a more careful investigation of the mechanisms underlying the repair of linguistic errors.

In other words, the application of SDT to the problem of language production (and its monitoring and control) is not just a methodological choice. Rather, the utilization of such a framework aims at making a much broader theoretical statement, namely, that language production, similar to many other cognitive systems, is a self-regulating performance-optimizing system, special in terms of its representations but not its operations (Nozari, 2018).

## Conclusion

In this study, we adopted the theoretical position that monitoring combines multiple sources of information (Nozari, 2020), the absence of each of which can decrease the quality of the monitoring signal and lead to poorer discrimination of errors from correct trials. We tested this prediction by applying the SDT framework to the detection and correction performance when some or all of visual information from the external channel had been removed. The prediction was borne out: removal of the visual signal affected both conscious detection and correction of typing errors, with corrections showing a greater dependence on such information. When part of visual information (positional information) was reinstated, the quality of signal increased, although not to the level of full information. Moreover, the application of SDT proved useful in interpreting how the system deals with the lower quality of signal when task demands are very different. Participants naturally shifted their criterion to avoid False Alarms when the cost of False Alarms were high, as in attempting corrections on already correct responses. When the cost was not so high, as in deciding whether an error had been committed or not, criterion was not shifted, showing flexible and adaptive use of information in optimizing behavior. Similarly, we found that participants' lower correction rates in the absence of visual information about letter identity reflected their accurate estimation regarding the insufficiency of the available information for undertaking repairs.

Collectively, these findings suggest that monitoring is a multi-channel adaptive mechanism that can be modeled as a decision-making process superimposed on a conflict detection framework. The goal of this system is not simply to detect as many errors as possible, but to optimize performance in accordance with task goals.

## Acknowledgments

This project was supported by the Therapeutic Cognitive Neurology Fund, to Johns Hopkins School of Medicine, Department of Neurology, Division of Cognitive Neurology.

## References

- Anders, R., & Alario, F.-X. (2019). Challenges to developing time-based signal detection models for word production. *Cognitive Neuropsychology*, 36(1–2), 85–88.  
<http://doi.org/10.1080/02643294.2019.1581603>
- Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, 82, 57–73.  
<http://doi.org/10.1016/j.cogpsych.2015.07.002>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459.  
<http://doi.org/10.3758/BF03193014>
- Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental Psychology* (2006), 68(7), 1268–1294.  
<http://doi.org/10.1080/17470218.2014.978877>
- Chomsky, N. (1968). *Language and mind*.
- Costa, A. (2019). A flexible criterion of response selection: When is it needed? Comment on Nozari and Hepner (2018). *Cognitive Neuropsychology*, 36(5–6), 208–211.  
<http://doi.org/10.1080/02643294.2018.1532404>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12. <http://doi.org/10.3758/s13428-014-0458-y>
- Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(2), 222–233. <http://doi.org/10.1037/0278-7393.10.2.222>
- Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word Production : Behavioral and Computational Considerations. In *The Oxford Handbook of Language Production* (pp. 88–104).
- Fodor, J. A. (1983). *The modularity of mind*. MIT press.
- Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning? *Cognition*, 176(June 2017), 107–130.  
<http://doi.org/10.1016/j.cognition.2018.03.009>
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594–621.
- Guenther, F. H. (2016). *Neural Control of Speech*. MIT Press.
- Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., ... Chan, P.

- (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, 48(3), 829–842.  
<http://doi.org/10.3758/s13428-015-0642-8>
- Hanley, J. R., Cortis, C., Budd, M. J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, 142, 36–47. <http://doi.org/10.1016/j.jecp.2015.09.008>
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113–157. <http://doi.org/10.1006/cogp.2000.0744>
- Hartsuiker, R. J., Pickering, M. J., & de Jong, N. H. (2005). Semantic and phonological context effects in speech error repair. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31(5), 921–32. <http://doi.org/10.1037/0278-7393.31.5.921>
- Hepner, C. R., Pinet, S., & Nozari, N. (2018). An enhanced model of gemination in spelling : Evidence from a large corpus of typing errors. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 487–492). Austin, TX: Cognitive Science Society.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–45. <http://doi.org/10.1038/nrn3158>
- Kalfaoğlu, Ç., & Stafford, T. (2014). Performance breakdown effects dissociate from error detection effects in typing. *Quarterly Journal of Experimental Psychology* (2006), 67(March 2015), 508–24. <http://doi.org/10.1080/17470218.2013.820762>
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25(4), 463–492. <http://doi.org/10.1080/02643290701674851>
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.  
<doi:10.18637/jss.v082.i13>
- Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. *Sentence Processing*, 281–294.
- Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech, Language, and Hearing Research*, 14(4), 677–709.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104. [http://doi.org/10.1016/0010-0277\(83\)90026-4](http://doi.org/10.1016/0010-0277(83)90026-4)
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, 25(6), 1198–1205. <http://doi.org/10.1177/0956797614529797>
- Logan, G. D., & Crump, M. J. C. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, 330(6004), 683–6.

- http://doi.org/10.1126/science.1190483
- Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, 125(4), 453–485. http://doi.org/10.1037/rev0000100
- Mahon, B. Z., & Navarrete, E. (2019). Adjudicating conflict in speech production—Do we need a central selection mechanism? *Cognitive Neuropsychology*, 36(5–6), 220–224. http://doi.org/10.1080/02643294.2019.1608171
- Makowski, (2018). The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, 3(22), 470. https://doi.org/10.21105/joss.00470
- Melinger, A., & Abdel Rahman, R. (2019). To re-tune or not to re-tune: Comments on the flexible criterion. *Cognitive Neuropsychology*, 36(5–6), 212–215. http://doi.org/10.1080/02643294.2018.1562886
- Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, 95, 19–35. http://doi.org/10.1016/j.jml.2017.01.007
- Nooteboom, S. G., & Quené, H. (2019). Temporal aspects of self-monitoring for speech errors. *Journal of Memory and Language*, 105(January 2018), 43–59. http://doi.org/10.1016/j.jml.2018.11.002
- Nooteboom, S. G., & Quené, H. (2020). Repairing speech errors: Competition as a source of repairs. *Journal of Memory and Language*, 111(November 2019), 104069. http://doi.org/10.1016/j.jml.2019.104069
- Nozari, N. (2020). A Comprehension- or a Production-Based Monitor? Response to Roelofs (2020). *Journal of Cognition*, 3(1), 1–46. http://doi.org/10.5334/joc.102
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1–33. http://doi.org/10.1016/j.cogpsych.2011.05.001
- Nozari, N., & Hepner, C. R. (2019a). To select or to wait? Response to the commentaries. *Cognitive Neuropsychology*, 36(5–6), 226–233. http://doi.org/10.1080/02643294.2019.1632280
- Nozari, N., & Hepner, C. R. (2019b). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, 36(5–6), 193–207. http://doi.org/10.1080/02643294.2018.1476335
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63(4), 541–559. http://doi.org/10.1016/j.jml.2010.08.001
- Nozari, N., Martin, C. D., & McCloskey, N. (2019). Is repairing speech errors an automatic or a controlled process? Insights from the relationship between error and repair probabilities in English and Spanish. *Language, Cognition and Neuroscience*, 34(9), 1230–1245. http://doi.org/10.1080/23273798.2019.1637007
- Nozari, N., & Novick, J. (2017). Monitoring and Control in Language Production. *Current*

- Directions in Psychological Science*, 26(5), 403–410.  
<http://doi.org/10.1177/0963721417702419>
- Nozari, N. (2018). How Special Is Language Production? Perspectives From Monitoring and Control. *Psychology of Learning and Motivation - Advances in Research and Theory*, 68(August), 179–213. <http://doi.org/10.1016/bs.plm.2018.08.006>
- Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, 30(2), 163–184. <http://doi.org/10.1023/A:1010377828778>
- Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive Neuropsychology*, 36(5–6), 216–219. <http://doi.org/10.1080/02643294.2019.1580189>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–47. <http://doi.org/10.1017/S0140525X12001495>
- Pinet, S., & Nozari, N. (2018). “Twisting fingers”: The case for interactivity in typed language production. *Psychonomic Bulletin and Review*, 25(4), 1449–1457. <http://doi.org/10.3758/s13423-018-1452-7>
- Pinet, S., & Nozari, N. (2020). Electrophysiological Correlates of Monitoring in Typing with and without Visual Feedback. *Journal of Cognitive Neuroscience*, 32(4), 603–620. [http://doi.org/10.1162/jocn\\_a\\_01500](http://doi.org/10.1162/jocn_a_01500)
- Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin and Review*, 23(6), 1898–1906. <http://doi.org/10.3758/s13423-016-1044-3>
- Postma, A., & Kolk, H. H. J. (1992). The Effects of Noise Masking and Required Accuracy on Speech Errors, Disfluencies, and Self-Repairs. *Journal of Speech, Language, and Hearing Research*, 35(3), 537–544. <http://doi.org/10.1044/jshr.3503.537>
- Postma, A., & Noordanus, C. (1996). Production and Detection of Speech Errors in Silent, Mouthed, Noise-Masked, and Normal Auditory Feedback Speech. *Language and Speech*, 39(4), 375–392. <http://doi.org/10.1177/002383099603900403>
- Postma, A., & Oomen, C. C. E. (2005). Critical issues in speech monitoring. In *Phonological Encoding and Monitoring in Normal and Pathological Speech* (pp. 157–166).
- Riès, S. K., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, 23(6), 1419–36. <http://doi.org/10.1162/jocn.2010.21467>
- Roelofs, A. (2020). Self-Monitoring in Speaking: In Defense of a Comprehension-Based Account. *Journal of Cognition*, 3(1), 1–13. <http://doi.org/10.5334/joc.61>
- Snyder, K. M., Logan, G. D., & Yamaguchi, M. (2015). Watch what you type: The role of visual feedback from the screen and hands in skilled typewriting. *Attention, Perception, & Psychophysics*, 77(1), 282–292. <http://doi.org/10.3758/s13414-014-0756-6>
- Tourville, J. a., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <http://doi.org/10.1080/01690960903498424>

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(2), 201–233.  
<http://doi.org/10.1037/xlm0000732>