The Perceptual Loop Theory Not Disconfirmed: A Reply to MacKay

WILLEM J. M. LEVELT

*Max Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands*

In his paper, MacKay reviews his Node Structure theory of error detection, but precedes it with a critical discussion of the Perceptual Loop theory of self-monitoring proposed in Levelt (1983, 1989). The present commentary is concerned with this latter critique and shows that there are more than casual problems with MacKay's argumentation. © 1992 Academic Press. Inc.

In Levelt (1989) I reviewed the existing theories of self-monitoring in speech production, and concluded:

> When we compare the editing account and the connectionist account, the most worrisome diagnosis is that, as they stand, both are very hard to disconfirm. (p. 476)

Since then, the situation has hardly changed. In his paper, MacKay (1992) outlines his connectionist Node Structure theory and cites some data with which it is compatible. But that is barely progress. What one would like to see is a concerted effort to disprove the theory's predictions in a truly Popperian fashion. However, neither MacKay nor others (including myself) have endeavored on this enterprise. As long as this is the situation, MacKay's theory is just one among other plausible accounts, no more, no less.

A bit more Popperian is MacKay's introductory discussion of the Perceptual Loop theory (PLT), my version of an editing account. MacKay tries to discount the theory on the basis of internal inconsistency and lack of empirical support. But his attack only shows how right I was when I wrote that even my own theory, as it stands, is hard to disconfirm. Let me go over MacKay's arguments.

In Levelt (1989) I cited the Lackner and Tuller findings as support for the existence of an internal and an external loop in self-monitoring. There is, in particular, the nontrivial result that speakers were faster in detecting their own speech errors when they listened to white noise (the masking condition) than when their overt speech was unmasked. In the masking condition speakers can use only their internal loop, and this loop is shorter than the external loop; this corresponds to shorter error detection latencies (and, of course, less detections). MacKay does not deny the force of this argument, but claims that other aspects of the Lackner and Tuller findings are hard to reconcile with PLT.

One aspect is that the detection of voicing errors (e.g., /di/ for /ti/) suffered

---

Commentary on D. G. Mackay (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition*, 1, 199–225.

much more from masking than the detection of place of articulation errors (e.g., /di/ for /gi/). I argued that the acoustic difference between voiced and unvoiced is much more substantial than the acoustic difference between places of articulation (such as alveolar versus velar). There is, however, no obvious reason to suppose that a similar state of affairs holds for internal speech (and it may be the other way around). Hence, masking (which obliterates the acoustic difference) will affect the detection of voicing more than the detection of place of articulation. MacKay argues that it is problematic to talk about the size of a phonetic difference in internal speech (I don't think that it is) and that my kind of argument requires a comparison between sizes of (internal) articulatory and (overt) acoustic features. The latter is incorrect. It only requires a measure of phonetic saliency *within* each of the two domains (internal and overt speech), not between them. If the sizes of voicing and place differences are the same in internal speech, but different in overt speech, then the Lackner and Tuller finding follows from PLT.

A second result that MacKay finds hard to explain by PLT is that voicing errors *occurred* more often than place of articulation errors. But PLT is a theory of error *monitoring*, not a theory of error *production*. Still, MacKay argues that the point is essential because "errors in general may have been detected faster with masking in Lackner and Tuller not because of reliance on an internal loop, but because of a reduced criterion for accuracy," i.e., a higher speed/accuracy ratio under masking. That may or may not be so, but it does not explain why this then leads to more voicing errors as opposed to articulation errors. Hence, these data are irrelevant for the evaluation of PLT.

MacKay then turns to other things that PLT does not or cannot explain. One example of the first kind is this: "If masking causes suppression of the external loop for error detection, how do speakers monitor and respond to loudness of the masking itself?" (i.e., by speaking louder). Not much of a problem here. PLT says that speakers *listen* while they speak. Hence they *hear* the noise and can respond to it. In PLT the decision of how to respond is a central one, not a mere automatism. The speaker can stop, speak a bit louder, remove the headphones, or start weeping.

Another example of the first kind is "how the perceptual system is able to detect deviations from linguistic standards." Here MacKay is right. I never developed a detailed account of how listeners detect linguistic ill-formedness, except for the detection of ill-formed compounds (Levelt, van Gent, Haans, & Meijers, 1977). And to the best of my knowledge, nobody else ever did.

An example of the second kind concerns the detection of lexical substitution errors (like *table* for *chair*). There is a "representational" problem here, according to MacKay, "because the perceptual system begins at best . . . with a phonetic representation of the output. However, lexical errors cannot be detected using a phonetic representation because nothing is phonetically wrong with words that are substituted in error." This is even a "logical inconsistency" in the theory, it is said. How does MacKay suppose a *listener* recognizes that a word is correctly used? There must be a "representational problem" and a "logical inconsistency" there as well because the listener surely begins with some kind of phonetic representation of the word. Still, if I point at a chair and say *table*, my hearer will

probably notice that I made an error. So, what's the problem? In PLT the speaker compares the meaning of a parsed word to the intended notion. If I intend to express the notion TABLE and then hear myself say *chair*, there will be a mismatch at the conceptual level, and I may become aware of it.

MacKay concludes this unusual "representational" argument as follows: "Such considerations suggest that instead of occurring at only one or two points or levels in the production process, editing must be a distributed or "everywhere" characteristic that permeates the entire output/perception process, from the highest level concepts to the lowest level phonetic units. To capture this distributed aspect of error detection, PLT requires as many loops as there are units involved in error." However, in PLT, monitoring does not take place in the production process at all, let alone at all of its levels. It is precisely the core proposition of PLT that monitoring is mediated by the speaker's language *comprehension* system. It would have been a good question to ask how the comprehension system spots trouble at different levels of processing (e.g., acoustic, phonetic, morphological, semantic, syntactic, pragmatic trouble). MacKay is apparently horrified by the idea that editing must be distributed over all levels of processing, and so am I. But that is exactly how his own Node Structure theory accounts for the detection of trouble in units of various levels or size.

A final shortcoming is quite cryptic: "if the same system processes other-produced and self-produced inputs, including errors, how do characteristics that fail to enter awareness when speech is produced correctly suddenly enter awareness when an error is produced?" Does MacKay mean that if other-produced and self-produced inputs were handled by *different* systems the answer would be obvious? I don't see the argument.

MacKay then turns to several predictions derived from PLT, sketching a bleak picture of their empirical support. The seven predictions and my comments are, in short:

(1) The detection of phonological errors should be faster than the detection of word errors.

This need not be so. We know that syllable detection is faster than phoneme detection and we know that there is a lexical effect in phoneme detection. Similar things may happen in error detection. It is not necessarily the case that lower level units can be attended to faster than higher level units. But whatever is the case, there are simply no data available that are relevant to this issue.

(2) Monitoring of self-produced speech and of other-produced errors should exhibit identical error tendencies.

This has been claimed before by MacKay (1987) and rejected by Levelt (1989, pp. 471–472). MacKay does not take the trouble to answer the criticism. It says, in short, that monitoring is task and context dependent, as any attentional process. The aims of monitoring are vastly different between a speaking task and a listening task. Hence, one should not expect identity of monitoring error tendencies.

(3) Speakers should sometimes produce pseudocorrections. If they misperceive a self-produced item, they should tend to correct it and produce the same item again.

That is a fine prediction, and I can remember making such pseudocorrections myself. But then I read "However, pseudocorrections have so far never been reported in cases of word meaning and word sound deafness, which is a problem for PLT." It is, I would say, quite a bit more problematic for error collectors; apparently nobody ever looked for them. A child can detect reduplications in spontaneous speech, but they are not the stuff that is collected by our industrious colleagues. If some of these reduplications turn out to be pseudocorrections, then that is problematic for the Node Structure theory.

(4) Overt speech errors should be easier to detect than the mental errors that occur during internal speech.

MacKay reports that such a difference was not found by Dell (1978, 1980), but that it did arise in Dell and Repka (in press). Hence, there is some support for the prediction.

(5) The lexical bias in speech errors (predicted by PLT) should also hold for blends.

There is overwhelming evidence for the existence of lexical bias in speech errors (see Levelt, 1989, for a review). There is no disagreement with MacKay that this is so nor that PLT can explain it. Lexical bias is also predicted by connectionist accounts of monitoring (cf. Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991). But according to MacKay there is a problem with blends. He discusses an unpublished study (by Collins and Ellis), to which I have no access; it does not find lexical bias in case of blends (such as *sotally* as a blend of *solely* and *totally*). That should then be a problem for PLT. But it isn't. According to PLT, what the speaker monitors for and intercepts is task and context dependent. Monitoring for lexical status is subject to selective attention, as was convincingly shown in an experiment by Baars, Motley, and MacKay (1975). Lexical bias could be induced or annulled, dependent on the contextual setting. It is never surprising when lexical bias does not arise. It does not arise, for instance, in del Viso et al.'s (1991) data, and these are not blends. And apparently it does not arise in one study on blends. There is no special status for blends, I would say, as far as lexical bias is concerned.

Later in the paper, MacKay considers whether PLT could be extendable to other output systems, such as typing and handwriting. I never claimed anything about typing or handwriting, and PLT was solely developed for self-monitoring in speech. But, for the case of argument, I will add these predictions here:

(6) Detection of correct responses should be faster than detection of errors.

This is, apparently, contradicted in research on typing. Why should PLT make that prediction? According to MacKay this is because there is much more experience with error-free output than with erroneous utput, so we should be faster on the error-free stuff. But this, again, ignores the attentional aspect of self-monitoring. If we would spend our limited attention exclusively on what we do correctly (because that is the default case), self-monitoring would betray its ecological function: to prevent error. PLT predicts that we will attend to trouble in the first place, and what you attend to, you do relatively fast.

(7) Error detection will improve depending on how much external feedback is provided or allowed in typing or handwriting.

MacKay mentions three studies that support the claim for relatively unskilled typists. The latter detect less errors when they cannot see the paper. One study finds the same for skilled typists when they reach for relatively unpracticed keys, but for practiced keys there is no difference. So there is good support for the prediction, except in the latter case. And there the explanation is obvious: external feedback becomes less necessary for practiced typists. They must be relying more on their internal feedback loop, and nothing in PLT forbids us to make the same claim for feedback in speech production.

So, what is the harvest? There were seven predictions. The first one does not necessarily hold, and there is no evidence for or against it. The second one is incorrect; it does not follow from PLT. There is no evidence for or against the third prediction. There is some empirical support for the fourth one, and strong support for the fifth. The sixth prediction does not follow from PLT, and the seventh prediction is supported in the majority of studies.

Question: has PLT been disconfirmed by MacKay?

REFERENCES

Baars, B. J., Motley, M. T. & MacKay, D. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–391.

Collins, A. F., & Ellis, A. (1991). *Blends, lexical bias and the concept of pre-articulatory editor in speech production. Unpublished paper*. University of Lancaster, England.

Dell, G. (1978). Slips of the mind. In M. Paradis (Ed.), *The Fourth Lacus Forum*. Columbia, SC: Hornbeam Press.

Dell, G. (1980). *Phonological and lexical encoding in speech production: An analysis of naturally occurring and experimentally elicited speech error*. Unpublished doctoral dissertation, University of Toronto.

Dell, G., & Repka, R. J. (in press). Errors in inner speech. In B. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum.

Del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161–186.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. van Gent, J. A. W. M., Haans, A. F. J., & Meijers, A. J. A. (1977). Grammaticality, paraphrase and imagery. In S. Greenbaum (Ed.), *Acceptability in language*. The Hague: Mouton.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–142.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. Berlin: Springer Verlag.

MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness & Cognition*, 1, 199–225.