CONSCIOUSNESS AND COGNITION 1, 231–240 (1992)

Problems with Popper: The Initial Goal Is to Develop Viable Theories, Not Disconfirm Them

DONALD G. MACKAY

Psychology Department, University of California, Los Angeles, California 90024

The Popperian epistemology underlying Levelt's (1992) commentary and other aspects of contemporary psychology has limited application and, in particular, does not apply to the creation or development of theory, the main goal of MacKay (1992a). This is relevant to Levelt's questions, "What has changed?" and "What is the harvest?": From a non-Popperian perspective, both changes and harvest are greater than Levelt's commentary would suggest and carry implications for the field at large.

Popper's (1959) call for a disconfirmation strategy has been taken very seriously in psychology. According to psychologists Ceci and Bronfenbrenner (1991, p. 28), disconfirmation constitutes the essence of all science, which they define as "a strategy of 'proof by disproof.'" Levelt (1992) is clearly operating within this Popperian framework when he laments that Perceptual Loop theory (PLT) and Node Structure theory (NST) are hard to disconfirm "as they stand," and when he concludes that "the situation has hardly changed" as a result of MacKay (1992a). Moreover, Levelt's "little-has-changed" verdict seems to be largely correct as applied to PLT and is indeed depressing if disconfirmation is considered the be-all of science.

However, there is much more to science than has been envisioned in the Popperian framework. In particular, Popper's disconfirmation strategy does not apply to *development of theory*, which was the main goal of MacKay (1992a). And with regard to development, much has changed in the NST as a result of MacKay (1992a). I begin with a summary of these changes, focussing especially on disconfirmability. I next examine the Popperian framework and its pitfalls in more detail and argue that the disconfirmation strategy applies at best to only a limited aspect of science. I then conclude with a non-Popperian answer to Levelt's (1992) question, "What is the harvest?"

WHAT HAS CHANGED?

Three main changes follow from MacKay (1992a). One is empirical in nature: New empirical relations have been pointed out, e.g., a special relation between errors and awareness, and new empirical possibilities have been suggested, e.g.,

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

1053-8100/92 $5.00

new types of error detection based on representations within connotative and visual concept systems.

The second change is that NST has become more detailed. For example, the theoretical conditions that are necessary and sufficient for awareness and error detection have been spelled out in NST, new variables such as the concept of "distance" have been illustrated in detail, and new explanations for old phenomena have been developed, e.g., the relation between speech rate and lexical bias in production errors.

The third change is that NST has become more general, applying now to error detection in speech, reading, writing, and typing. These extensions have suggested new methodological paradigms for investigating error detection and new insights into classical paradigms such as the lexical decision task.

These three areas of change have enabled NST to make new, nonobvious and disconfirmable predictions, e.g., that error detection will decrease with "distance" and that errors involving highly practiced nodes will be easier to detect than errors involving less practiced nodes, all other factors being equal (*aofbe*). True, no concerted efforts have yet been made to empirically disconfirm these predictions. However, developing clear theoretical predictions can be useful even though these predictions remain untested or even untestable in practice for many years, witness, e.g., gravitational field theory in the period 1904–1918.

POPPER'S FRAMEWORK AND ITS LIMITATIONS

What is Popper's (1959) framework and is it universally useful across all aspects of science? Popper's framework contains five central assumptions expressed with varying degrees of elaboration and explicitness. I will first spell out these assumptions and then argue that some are false and others are useful only within limited contexts. In particular, I will argue that Popper's framework is not useful for achieving MacKay's (1992a) goal of theory development.

Popperian Assumption 1: Reject Disconfirmed Theories

Popper's main assumption was that theories should be rejected as soon as disconfirming evidence appears. However, a strong case can be made that Assumption 1 is not what actually happens even in the best of science. As Kuhn (1970) points out, scientists often "maintain a theory as a whole even though it has been falsified by single experiments or other observations . . . Theories are only rejected when all their important propositions must be revised and/or a new and better theory arises."

More importantly, a strong case can be made that Assumption 1 is not what *should* happen in the best of science. If theories were rejected as soon as disconfirming evidence appeared, no new theories could be developed because one or more pieces of existing evidence invariably contradict new theories (see MacKay, 1992b, for examples). New theoretical ideas such as NST and PLT must be protected against the premature dismissal that Popper's disconfirmation strategy demands.

Even in the case of a well-developed, well-established, and well-supported

theory, contradictory empirical phenomena do not and should not discredit or bring disapproval on the theory because such theories should not be overthrown lightly, and because falsifying a prediction is not a simple or straightforward affair (see Quine, 1961, and Duhem, 1953). Predictions from a theory cannot be tested in isolation from the network of (often implicit) theoretical assumptions concerning the situation of test. Moreover, theories commonly fail not through some fault of the theory per se, but because, nature being complex, the situation of test violates some *aofbe* assumptions of the theory. Specifying these *aofbe* assumptions is therefore an important precursor to testing a theory and to evaluating tests of a theory.

Popperian Assumption 2: Disconfirmation Is Central to All of Science

Under Popper's second assumption, the process of rejecting disconfirmed theories (under Assumption 1) is central to all aspects of science. Assumption 2 is questionable because eliminating incorrect (disconfirmed) theories in favor of correct (undisconfirmed) ones is only one of many goals in science. Another goal is practical applications for which even incorrect theories are invaluable (see MacKay, 1988). An equally important goal is to create or develop viable theories, and here the disconfirmation strategy is simply irrelevant. Large scale disconfirmation might challenge or even damage a well-developed theory, but it cannot create a new one.

Moreover, Assumption 2 can be counterproductive even with respect to the goal of disconfirming theories. There are many reasons why theories may be hard to disconfirm "as they stand." One is that the theory and its competitors are underdeveloped and neither clear enough nor specific enough to enable conclusive empirical test. Far from being "worrisome," however, this situation is actually typical of newly proposed theories and can be altered by further developing the theories to a point where they *are* clear and specific enough to discriminate via empirical test. When theories are first proposed, demanding the possibility of immediate test is counterproductive (see also, e.g., Feyerabend, 1988; Carnap, 1966).

Popperian Assumption 3: Only Disconfirmation Counts in Theory Evaluation

Under Popper's third assumption, survival following concerted attempts at empirical disconfirmation is the only proper criterion for evaluating theories. However, all theories face varying degrees of disconfirming evidence and can in general be considered false (Robinson, 1984), so that Popper's followers have adopted a curious corollary to Assumptions 1–3. Under this corollary, the theory with the fewest exceptions or contradictory observations should be accepted. However, theory development is not an exercise in empirical democracy. Theories must explain the how and why of well-established exceptions as well as the rules.

This point is relevant to Levelt's (1992) comments on the relation between error detection and amount of external feedback in PLT. According to Levelt, the generalization that error detection will improve as more external feedback is

provided or allowed in speech, typing, or handwriting is "supported by the majority of studies." The fact that external feedback becomes irrelevant to error detection when a skill becomes highly practiced or proficient can apparently be disregarded as a "minority vote" (exception or aberration). However, this particular well-established exception demands explanation as the prototypical case or focus of interest for many, including Levelt (1989): All of adult speech production constitutes a proficient skill.

Returning to the core of Assumption 3, testing a well-developed theory is clearly useful and desirable, but disconfirmation and more generally, compatibility with observation, is just one of many other factors by which theories in general are and should be evaluated (see MacKay, 1992b). These other factors include coherence, scope, diversity of explanation, specificity, heuristic value, and parsimony. Thus, theories that are coherent or "make sense" are to be preferred, *aofbe*; Theories that have larger scope or summarize a broader range of observations are to be preferred, *aofbe*; Theories that explain a wide range of diverse or disconnected empirical laws are to be preferred over theories that explain a narrow or closely connected set of empirical laws, *aofbe*; Theories with greater specificity of mechnism or "depth of penetration" (Royce, 1988) are to be preferred, again, *aofbe*; Theories that stimulate advances in knowledge are to be preferred, *aofbe*, quite independently of whether the new knowledge is compatible or incompatible with existing theories; Theories that are elegant and simple are to be preferred, *aofbe*, and parsimony is one of the reasons why I originally chose to analyze PLT rather than some other theory for comparison with NST: PLT seemed to explain a range of phenomena on the basis of a very simple mechanism, a loop that links mechanisms for inner speech to mechanisms that are needed anyway for perceiving overt speech.

The mere existence of these "other factors" in theory evaluation contradicts Assumption 3. Moreover, a strong case can be made that these "other factors" are more important than (non)disconfirmation. As Brandt (1984) points out, theorists generally revise or reject a theory not because it fails some test or proves difficult or impossible to test but because of incoherence: the theory no longer makes sense to them. Moreover, the importance of these "other factors" for theory evaluation helps explain why theories are valued and used long after they have proven inadequate or insufficient for explaining available facts (Kuhn, 1970), a phenomenon that is paradoxical under Popper's third assumption.

Popperian Assumption 4: All Undisconfirmed Theories Are Equal

Under Popper's fourth assumption, all available theories that apply to some set of phenomena should be considered equal until efforts to disprove one of them fail. Levelt (1992) seems to adopt Assumption 4 when he suggests that unless a concerted effort is undertaken to disprove the NST of error detection, it remains "just one among other plausible accounts, no more, no less." However, if theories are evaluated on many other dimensions besides their state of disconfiguration or compatibility with observation, as the above analysis suggests, then Assumption 4 is false.

As a corollary to Assumption 4, Popper (1959) argued that disconfirmation is necessary for proposing new theories or replacing one theory with another: Theories should be accepted *until* they have been disconfirmed, according to Popper. However, this corollary is also open to question because creating new or alternative theories is often necessary in order to discover, seek, or bring to light new data that would delimit, challenge, or disconfirm an established theory (see MacKay, 1992b, for relevant examples).

Popperian Assumption 5: Verificationism Is Unscientific

Under Popper's fifth assumption, attempts to verify a theory or demonstrate its compatibility with particular empirical situations or findings are counterproductive, and Levelt (1992) seems to share this dislike of verificationism (which he describes as "barely progress"). However, when it comes to creating or developing theories rather than rejecting them, verificationism is not just acceptable but useful as an initial strategy: Theoretical development often proceeds a long way by examining whether and how a theory works in a variety of particular instances, including hypothetical or imaginary instances (Gedanken experiments; see MacKay, 1992b). Applications to instances help to develop a theory by challenging its viability, clarity, and generality.

Does the Popperian Framework Apply to MacKay (1992a)?

Insofar as Levelt (1992) has accepted the Popperian framework, it is understandable that he interpreted MacKay (1992a) as an attempt to disconfirm or discount PLT. However, Popper's framework, as outlined above, is irrelevant to the main thrust of MacKay (1992a), which was not to disconfirm theories (PLT and NST in particular), but to challenge them and further their development. Thus, lacking clear predictions, NST and PLT *were* hard to disconfirm in 1989. But whereas Levelt found this situation "worrisome," MacKay did not because his goal was to further develop these theories to a point where they are clear and specific enough to be discriminated via empirical test.

WHAT IS THE HARVEST?

At the end of his commentary, Levelt (1992) raises the question, "What is the harvest?" With regard to the goal of disconfirming PLT, not only is the harvest poor, as Levelt rhetorically suggests, but the field has yet to be planted. With regard to other scientific goals, however, the harvest seems richer. I discuss three such areas of harvest below.

Theoretical Development

With respect to PLT, MacKay's (1992a) main point was that PLT is an interesting and important theory that suffers from underdevelopment. To further illustrate this need for development, consider Levelt's (1992) reference to the perceptual precedence of higher level units, the fact that higher level units such as words and syllables can be detected faster than lower level units such as phonemes.

From a theoretical perspective, the issue is not whether PLT or any other theory can escape possible disconfirmation because of perceptual precedence, as Levelt seems to suggest. Rather, perceptual precedence is an empirical phenomenon that requires theoretical explanation, and the issue is how (by what internally coherent mechanisms) can PLT or any other theory achieve perceptual precedence for higher level units, and what role do these mechanisms play in error detection (see, e.g., the account in NST; MacKay, 1987, p. 67). As Levelt (1992) correctly notes, no currently available data bear on this issue. However, this fact is irrelevant from a theoretical perspective: Theoretical development can and should proceed in advance of empirical evidence (see MacKay, 1988). After all, theories are meant to predict, which implies going beyond the available data.

Nonetheless, PLT is not entirely without harvest from MacKay's (1992a) challenges: By addressing these challenges, Levelt (1992) has developed PLT to some extent. Levelt's response to the "representational argument" constitutes one area of development. According to Levelt, "the speaker compares the meaning of a parsed word to the intended notion." However, the issue that MacKay had raised was how or by what detailed theoretical mechanisms this comparison of production with perception processes occurs, and Levelt seems to accept MacKay's proposition that "to capture this distributed aspect of error detection, PLT requires as many loops or connections between perception and production units as there are units involved in error."

Multifold perception-to-production loops are certainly not a "horrifying" direction of development for PLT. As Levelt (1992) points out, multifold perception-to-production loops give PLT a distributed characteristic that resembles the NST. However, adding so many loops reduces the simplicity of PLT, one of its original attractions.

Nor are multifold perception-to-production loops an "obvious" direction of development for PLT. I had expected a quite different solution to the representational problem within PLT, namely to introduce into PLT's perceptual system a set of top-down connections that originate in the production system. These top-down connections would have little impact on parsimony because they are needed for explaining many other aspects of perception (see MacKay, 1987, pp. 35–36). Again, however, assuming top-down connections within the perceptual system would make PLT more similar to NST.

A more problematic developmental harvest for PLT concerns the Lombard effect, the fact that speakers increase their output amplitude when background masking or white noise is increased in loudness (Fairbanks, 1954). As MacKay (1992a) notes, Lombard effects are hard to understand if, as PLT originally assumed, masking causes suppression of the auditory pathway or external loop. To solve this problem, Levelt (1992) apparently assumes that masking only *partially* suppresses or attenuates use of the external loop: Speakers can hear the attenuated masking noise and voluntarily respond to its increased loudness in the same way that they might decide to remove their headphones (or not).

However, this new hypothesis is itself problematic. The increased loudness of speech following a sudden increase in white noise is reflex-like and does not resemble voluntary responses such as removing headphones. For example, in-

structing subjects to keep the level of their vocal output constant despite increases in background noise does not diminish the Lombard effect (Siegel & Pick, 1974). Perhaps the partial suppression of the external loop that results from masking in PLT is enough to interfere with processing of *phonological* features such as voicing, but not *acoustic* features such as loudness (for some reason that requires specification in PLT). Or perhaps loudness is simply more difficult to suppress than voicing or place of articulation (for some reason that requires specification in PLT). In either case, further development of PLT is required to fully capture the Lombard effect.

Clarification of Challenges for PLT

A minor harvest of the present exchange is that it has enabled clarification of MacKay's (1992a) challenges to PLT. One clarification concerns the differential effect of masking on detection of errors in voicing versus place of articulation in Lackner and Tuller's (1979) data. MacKay was raising three questions: whether voicing (unlike place of articulation) involves a small production difference that translates into large acoustic effects, as Levelt (1989, p. 472) assumed; whether such a "size comparison" makes sense in principle, especially in the absence of a theory that explains how a small egg (muscle movement commands for voicing and place of articulation) turns into a large chicken (acoustic aspects of voicing and place of articulation); and whether, if such a theory existed, it would have a central place for the notion of size (just as the notion of size plays no central role in the double helix theory that explains how small eggs turn into large chickens).

The issue was not whether talking about the size of a phonetic difference in internal speech is problematic (which it is, especially if *phonetic* units play no role in inner speech; see MacKay, 1987, p. 24). Nor was the issue whether the size of acoustic features can be compared for internal versus overt speech (which they cannot because no acoustics arises from inner speech). However, Levelt (1992) has clarified that the term *salience* rather than *size* more closely matches his original intent, so that the initial issue is this: Does producing voiced versus unvoiced speech sounds involve a nonsalient production difference that translates into a salient acoustic difference (unlike producing front versus back speech sounds)? The answer to this question will depend in part on what is meant by a nonsalient production difference: Although salience has a clear and well-established meaning for *perception*, the same is not true for production, especially the normally unconscious production of phonetics.

Another minor clarification concerns the relation between masking noise and speed–accuracy trade-off. MacKay's (1992a) idea was that hearing white noise may have induced subjects in Lackner and Tuller (1979) to allow a higher error rate (adopt a lower criterion for accuracy) both in producing the syllable strings and in analyzing the syllables that they produced. (The idea that accuracy in perception and production should be coupled or respond in the same way to factors such as masking makes good sense in the NST where perception and production engage identical phonological nodes).

Now, because accuracy (errors) can trade off with speed in both perception

and production (see MacKay, 1987, pp. 91–92), a lower criterion for accuracy would have enabled a faster response rate, which may explain why response times were faster with than without masking noise in Lackner and Tuller. Moreover, because this idea also explains why more production errors occurred with than without masking in Lackner and Tuller, the speed–accuracy hypothesis is more parsimonious than Levelt's (1989, pp. 471–472) hypothesis that masking causes greater reliance on an internal loop. Finally, if perception and production involve the same phonological nodes, as assumed in NST, the fact that masking interferes more with correct perception of voicing than articulatory place (within the range of signal-to-noise ratios from 0 to −18 db; Miller & Nicely, 1955) would explain Lackner and Tuller's remaining phenomena; that in the masking condition, voicing was more problematic than place of articulation in both production (relatively more voicing errors occurred) and perception (relatively fewer voicing errors were detected).

A final clarification concerns MacKay's (1992a) statement that "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?" The idea was not that different systems handle other-produced versus self-produced inputs, but rather that the single perceptual system in PLT should process phonetic aspects of self-produced inputs identically, for errors and nonerrors. Thus, the as-yet-unaddressed challenge for PLT is this: If phonetic aspects of correct inputs, e.g., the duration of fricatives, never enter awareness when perceiving speech (including one's own), then how do these aspects enter awareness when speakers produce a phonetic error, e.g., slurring of a fricative?

Clarification of PLT

Clarification of PLT represents a final area of harvest: Levelt's (1992) commentary has furthered our understanding of how PLT works by specifying some *aofbe* assumptions of his theory, an important precursor to viable tests of PLT. Levelt's references to attention and its vicissitudes illustrate some of these *aofbe* assumptions. However, PLT requires further clarification on this and several other points before disconfirmation becomes feasible. Levelt's concept of attention is so overworked and so underspecified as to threaten further development of the theory. For example, if error detection exhibits different patterns in self-produced versus other-produced speech, this is unproblematic for PLT because attention may be allocated differently to errors when speaking versus listening (although the what and how of attentional allocation to errors remain unspecified). If lexical bias does not hold for blends, this too is unproblematic for PLT because the contextual setting can eliminate lexical bias by altering the attention allocated to lexical status (left unspecified is how and why attention is allocated differently for blends,[1] Baars, Motley, & MacKay, 1975, notwithstanding). If detection of cor-

[1] In this regard, Collins and Ellis (1991) seems quite different from Del Viso, Igoa, and Garci-Albea (1991). Collins and Ellis made concerted efforts to obtain a lexical bias for blends, using many analytic

rect responses is faster than detection of errors, this too is unproblematic for PLT because "we don't spend our limited attention only on what we do correctly, but attend to trouble in the first place." To develop as a theory, PLT must specify *how or by what mechanisms* we detect trouble in the first place and then, once detected, how we spend our limited attention on trouble rather than on correct output. Achieving this will constitute a rich harvest indeed for PLT.

CONCLUSION

The Popperian epistemology adopted by Levelt (1992) and many other psychologists has hidden pitfalls, and unrestricted application of this epistemology may be hindering the development of theories and distorting their evaluation in psychology. What seems needed is a new "theoretical" epistemology (see MacKay, 1992b) to supplement the "empirical" epistemology within which Popper's contribution is one rather defective or overgeneralized component. Whereas the empirical epistemology has proven useful for developing reliable observations over the past 75 years in psychology, a theoretical epistemology is needed for developing viable theories. Attaining this theoretical epistemology will bring a rich harvest for psychology at large (see MacKay, 1992b).

ACKNOWLEDGMENTS

This research was supported in part by NIA Grant AG09755-02. The author thanks Dr. D. Burke for comments on an earlier version of this reply.

REFERENCES

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

Brandt, L. W. (1984). Logic and psycho-logic of science. *Annals of Theoretical Psychology*, **2**, 203–210.

Carnap, R. (1966). *Philosophical foundations of physics*. New York: Basic Books.

Ceci, S. J., & Bronfenbrenner, U. (1991). On the demise of everyday memory. *American Psychologist*, **44**, 27–33.

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

Del Viso, S., Igoa, J. M., & Garci-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.

Duhem, P. (1953). *The aim and structure of physical theory* (P. Weiner, trans.). Princeton, NJ: Princeton Univ. Press.

---

procedures and several corpuses of blends. However, Del Viso et al. examined a relatively small number of phonological errors (454) in Spanish and obtained a small but nonsignificant effect in the right direction (37% lexical outcomes compared to 33 percent by chance). We already know that lexical biases are small, and nonsignificance can be attributed to an infinity of other variables in the context of test. An interesting possible "other variable" in the present case (G. Dell, personal communication) is speech rate, a factor known to influence lexical bias: Spanish may exhibit reduced lexical bias because it is spoken more rapidly in phonemes per second than, say, English.

Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Research*, **19**, 133–139.

Feyerabend, P. K. (1988). *Against method*. London: Verso.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: Univ. of Chicago Press.

Lackner, J. R., & Tuller, B. H. (1979). The role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 281–294). Hillsdale, NJ: Erlbaum.

Levelt, W. J. M. (1992). The perceptual loop theory not disconfirmed: A reply to MacKay. *Consciousness & Cognition*, **1**, 226–230.

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

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

MacKay, D. G. (1992b). The theoretical epistemology: A new perspective on some long-standing methodological issues in psychology. In G. Keren & C. Lewis (Eds.), *Methodological and Quantitative issues in the analysis of psychological data* (pp. 229–255). Hillsdale, NJ: Erlbaum.

MacKay, D. G. (1988). Under what conditions can theoretical psychology survive and prosper?: Integrating the rational and empirical epistemologies. *Psychological Review*, **95**, 559–565.

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

Miller, G. A., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338–352.

Popper, K. R. (1959). *The logic of discovery*. London: Hutchinson.

Quine, W. V. O. (1961). Two dogmas of empiricism. In W. V. O. Quine (Ed.), *From a logical point of view* (2nd ed., pp. 20–46). New York: Harper & Row.

Siegel, G. M., Pick, H. L. (1974). Auditory feedback in the regulation of voice. *Journal of the Acoustical Society of America*, **56**, 1618–1624.

Robinson, H. J. (1984). A theorist's philosophy of science. *Physics Today*, **37**, 24–32.

Royce, J. R. (1988). The implications of differential theory appraisal and the context of discovery for advancing theory in psychology. In W. J. Baker, L. P. Mos, H. V. Rappard, & H. J. Stam (Eds.), *Recent trends in theoretical psychology* (pp. 59–64). Berlin: Springer-Verlag.