AGING SV's – WE HAVE SOLUTIONS

James L. Farrell, *VIGIL Inc*.

Institute of Navigation GNSS 2009
Savannah GA

BIOGRAPHICAL

James L. Farrell (Ph.D., U. of MD `67; former ION Air Nav representative, senior member of IEEE, former AIAA local board member, member of τβπ, ηκν, πμϵ) is a registered professional engineering consultant in Maryland. Work includes teaching at Marquette and UCLA, two years each at Honeywell and Bendix-Pacific, plus 31 years at Westinghouse in design/simulation/ validation of estimation algorithms for navigation and tracking [*e.g.*, F16 AFTI, B1 radar, SDI; fire control system design, test data generation for bench validation; INS update and transfer alignment algorithm design; program development for USAF-WPAFB (F/C evaluation) and NASA (estimation of satellite attitude, orbit, deformation); missile guidance optimization, MLE boundaries] plus digital communication design (sync, carrier tracking, decode). He is author of *GNSS Aided Navigation and Tracking* (2007; NavtechGPS dist.), *Integrated Aircraft Navigation* (Academic Press, 1976; now in paperback after five printings), and of various columns plus over 80 journal and conference manuscripts. Active in RTCA for several years, he served as co-chairman of the GPS Integrity Fault Detection and Isolation Working Group within SC-159.

ABSTRACT

SV aging prospects, recently receiving attention in the navigation community, are not viewed with equal concern by all. With the stakes so high, however, even a small likelihood of reduced capabilities calls for a prudent assessment of resources. This paper identifies means of maintaining operational capability with reduced information. Many of those means are familiar but some are quite recent – and thus currently unused.

Discussion involving significant modification of operations is incomplete without mentioning certification. Addressing that subject raises some severe limitations of existing procedures that are unfortunate/inconvenient but nevertheless essential to confront. Even if existing institutional constraints preclude changing near-term approval requirements, the industry needs to become better prepared for unforeseen events that could affect the next generation. Consequently the proposed paper includes both technical solutions and some basic considerations involving validation.

BACKGROUND

GPS is by far the greatest system for both navigation and timing that this planet has ever seen – recognition of that is essentially universal. GPS lives up to expectations, brilliantly performing as advertised. Even that best-ever performance must have (and does have) tolerance for occasional error; examples of that, though rare, are well documented. To live with less-than-perfect performance, the industry relies on integrity testing (wherein comparison checks using extra satellites can detect inconsistencies and exclude questionable data). Methods to perform those tests are firmly established and supported by documented results.

Nevertheless it is universally recognized that GNSS, even with existing Fault Detection and Isolation or Exclusion (FDI/FDE), is still not perfect. Consequently, ramifications of growing dependence on GPS have been attracting more attention. It is insufficient to admit that no system is perfect; clear perspective calls for some quantitative measure. The overall subject can subdivide into general areas involving the likelihood of (1) reduced availability and (2) reduced "dependability" (terminology used here to include integrity plus the verification thereof, plus more – *e.g.*, backup). Although the first item is the main topic addressed here, the second presents itself as an unavoidable relevant issue – to such an extent that it is difficult to keep the topics separate. Still, because the second item is not dominant in the present context, much of the coverage for verification is placed in an Appendix. Ahead of that it suffices to note here that reasons for caution are very real. A dramatic example: despite wide and fully earned acclaim for the excellent 2001 Volpe report [1], commitment to a key means of backup for GPS remains unclear at the time of this writing.

Concerns for both near-term and long-term (next generation) future exhibit a disconnect. Prospective limitations (*e.g.*, reduced information – whether due to satellite aging [2], increasingly hostile electromagnetic environments, or other unforeseen events) can block satisfaction of beneficial capabilities (*e.g.*, to accommodate a desire for closer spacing with increased traffic growth). Possibility of a shortfall calls for a review of

- existing methods and procedures, followed by
- possible means for closing the gap.

Those two considerations form the main sections that follow.

PRESENT METHODS (WHAT WE DO NOW)

Today's designs are configured to satisfy specifications that heavily emphasize accurate knowledge of instantaneous position. Ramifications can be summarized as in the subsections below.

Requirements for Full Fix + RAIM

When each data vector must be a self-sufficient source of instantaneous position, a requirement arises for enough satellite sightline directions with geometric spread at all times. That interdependence is magnified when more satellites are added to provide FDI/FDE, requiring every subset of four within the enlarged group to support the requisite geometry. With this "all-or-nothing" posture, data lapses are a major stumbling block. A data gap that is only partial becomes equated to "Loss of GPS."

Position-Oriented Approach

Especially at high speeds (*e.g.*, in flight), instantaneous position is highly perishable. With little or no emphasis placed on accurate dynamics (beginning with velocity), demand for continuously accurate instantaneous position is highly dependent on abundant data. That abundance includes sufficiently high data rates, since latency becomes a significant liability without usage of a dynamic file

Carrier Phase (Classical)

Successful usage of carrier phase information is decades old and widely documented. Although ambiguity resolution is not required in all carrier phase applications, requirements for cycle slip detection are quite common. More common yet – in fact, virtually ubiquitous – is the need to maintain phase continuity via a carrier track loop. When those needs are satisfied, sub-wavelength instantaneous position is obtainable. Challenges involved, however, have produced among users a wide variation in perception of value. Some negative perceptions have arisen due to "cutting corners" in formation of carrier phase, or merely settling for deltarange, by some receivers. Also a cycle slip, even if only rarely overlooked, can be catastrophic in some operations.

Validation Process: Very Imperfect

As already noted, verification is not the main topic of this paper – but the issue is inescapable. This passage briefly mentions shortcomings of existing procedures by citing

- hard evidence of certification improperly bestowed [3]
- severe limitations of GO/NO-GO criteria

and leaving discussions of additional examples plus their extensive ramifications for the Appendix.

BASIC CHANGES (WHAT WE'LL NEED TO DO)

Extremely powerful and versatile means to improve performance have been available for a very long time. Out of sheer necessity, ancient mariners fully exploited every opportunity to integrate partial data. Kalman's original paper [4], appearing a half-century ago., formalized a way to do that optimally. While Kalman estimation is commonly used today, its effective reach is almost invariably limited to data resident within each proprietary "box" of equipment.

Of course these facts, plus the resulting penalties in every facet of performance, are likewise well known and extensively documented. *So are the resources* for providing centrally processed solutions for data from

- every source of information available
- any combination of sources
- any subset that may exclude any sensor or group
- any individual source in a federated configuration.

Every conceivable choice from among these solutions can be made concurrently available – ***note the inherent backup***. All this capability is sacrificed by continued usage of

- interfaces chosen poorly or from outdated standards [5-8],
- undue consolidation within isolated equipment packaging,
- overextended proprietary rights, and
- limited – and demonstrably flawed – validation methods.

Explanations follow.

Relaxation of Requirements for Data Snapshots

An immediate explosion of benefits can follow from acceptance of partial information. Countless examples could be cited, but a few obvious ones are sufficient here:

- First, within GPS or GNSS, not all SVs would be simultaneously affected by scintillation (*e.g.*, ionospheric disturbance effects vary with both location and time). A similar case is made for multipath. ***Data from some SVs could be rejected, by decisions made external to a receiver, without forcing rejection of all***.
- Merging of GNSS data with information from other sources, including those external to any GNSS receiver (DME, Loran/eLoran • • •), offers another enormous improvement in multiple criteria for performance.

Again, these facts are so firmly established that there is no need to dwell on the myriad ways they can be exploited. Instead, broad statements will suffice here to

- <u>express the underlying goal</u>: Take advantage of available information that is not currently used
- <u>explain why the stated goal is not satisfied already</u>: GPS availability *thus far* has been more than satisfactory to a multitude of users (though that could change)
- <u>prescribe corrective strategies</u>: Many of those advocated, though appearing in various conference proceedings and journals, are cited from one reference [9] for compactness.

Availability Enhancements – a Start

For about two decades the industry was effectively guided by a strong preference for the trait, previously noted, whereby every data refresh event was self-sufficient. A major reason offered to support that preference was protection against gradual veering – a snapshot sequence is less sensitive than a continuously evolving path estimate. The cost, of course, is forfeit of benefits conferred by the sequence's history. More recently a middleground was sought to mitigate the resulting loss; "subfilters" used as much new data as possible while making some use of knowledge from an estimator's covariance matrix. An excellent post-SA example of that practice is given in [10].

This writer promptly endorsed that approach and sought to carry it to the limit. A "single-measurement RAIM" resulted (pages 121-126 of [9]) which, as its description implies, offers an independent integrity test for each separate observation. Despite its rigorous derivation, the technique is quite simple – even intuitive – in practice. Furthermore it bridges a gap that formerly separated integrity test from optimal estimation [§] while also having some significant advantages over conventional RAIM:

- separation translates to independence from other satellites – and therefore from geometry (effective DOP of unity)
- ability to use different error variances for different observations (*e.g.*, with nonuniform signal strengths etc.)

With this discussion we have clearly left the realm of well known subjects with self-evident prescriptions. Much of what follows will likewise fall into the category of relatively obscure methods.

Beyond a Position-Oriented Approach

A time history of GNSS observations, with or without an IMU, inherently carries dynamic information. A file with observational history from multiple sources of course enables the aforementioned explosion of benefits. Beyond the obvious immediate offerings, *i.e.*,

- closing of data lapses via information sharing
- intrinsic backup with automatic activation
- vast reduction of latency effects (*e.g.*, from 200 meters to less than 1-m at 400 kt, after 1-sec with easily obtainable velocity accuracy below 1 m/sec)
- formation of 1σ projected future error (within reason)

there are, once again, some lesser known techniques (including a few that are virtually nonexistent in operation at the time of this writing). With GNSS the full potential of dynamics calls for a revisit of carrier phase.

[§] With nonuniform and/or correlated measurement errors, correct error variance (Eq. 6.6 of [9]) and Kalman gain (Eq. 6.72 of [9]) come from factoring *weighted* – not unweighted – measurement sensitivity matrices for parity.

Carrier Phase (Recent Developments)

Rather than pursuit of unnecessary sub-wavelength fixes for aircraft (*e.g.*, with 20-m wing span moving at 400-kt), the true value of carrier phase in flight lies in enhanced dependability. [11] Sequential changes in carrier phase over 1-sec provide excellent dynamics information, again with or without an IMU (as shown in [9] by rigorous investigation with sequential correlations issues resolved – and supported by flight validation, also fully documented).

Recognition of the opportunity just identified led to the concept of segmentation, whereby position is determined separately from dynamics. Carrier phase sequential changes with ambiguities unresolved can provide precise (1-cm/sec RMS with IMU; decimeter/sec without) streaming velocity independent of position. Dead reckoning then provides *a priori* position correctible by pseudoranges. One particular advantage of this scheme is subtle – with 1-sec phase change propagation effects generally at 1-cm or less (pages 88-90 of [9]) no mask is needed. The geometry benefit is obvious and, also worthy of mention, flight experience also verified it : page 180 of [9] recalls rejecting several consecutive pseudoranges but not carrier phase changes from that same SV (it was at extremely low elevation). Immediately that raises another noteworthy segmentation characteristic: the previously mentioned single-measurement integrity testing is applicable to each carrier phase sequential change and to each pseudorange, separately and independently.

Capabilities discussed in this section are untapped in essentially all operational systems. Still another opportunity can be added – ability to sustain operation even if *every* SV has repetitive data gaps. This last advantage is best exploited with receivers described in the next section.

FFT-Based Processing Approach

Correlators and track loops in GNSS receivers can be replaced [12]; only a brief review and a list of basic advantages (page 146 of [9]) will be discussed here. The theory is age-old: multiplication in the frequency domain corresponds to convolution in time (and vice-versa). Thus a term-by-term product of a digitized receiver input's FFT with the reference pattern's FFT can, after an inverse FFT, provide outputs equivalent to full sets of correlator responses. Today's processing and A/D capabilities offer feasibility.

In addition to reduced vulnerability to jamming (not covered here), advantages are:

- access to all cells (not only a track loop's subset)
- *guaranteed* access (stability is not conditional)
- linear phase-*vs*-frequency; no phase distortion

Features from the preceding section, combined with these traits, offer extreme robustness.

Extension to Surveillance

The practice of transmitting responses to *r-f* interrogations has, for *many* decades, been quite vulnerable to overload (garble; one user's information is everyone else's interference). An important paper [13] described the unsurprisingly poor performance during the first Gulf war. – and proceeded to identify a remedy: squitters with separate assigned time slots. Immediately a sea change in capability offers every participant an opportunity to track every other participant. With no interrogations, garble would disappear.

This dramatic increase in capacity, furthermore, has been successfully demonstrated with the use of an existing communication link and existing airborne equipment (GPS receivers and Mode S squitters [14]). Subsequently this writer enthusiastically advocated adoption of the technique with one fundamental modification: Replace the data bits of the transmitted messages with measurements instead of coordinates [15]. Additional details such as

- the small shift in time (used in [14] to simplify time tagging without taxing link capacity) and
- adjustment to recompute measurements that would have been observed at the cg, to mitigate rotation effects

are briefly discussed on pages 143-144 of [9]. The main purpose of the recommendation, however, remains as stated – a host of major improvements listed at the beginning of [15], expanded even further in [16], and summarized on page 187 of [9]. The list of benefits is compelling.

CONCLUSIONS

Capability and dependability of navigation and surveillance can be enormously increased. The key lies not in new inventions nor provisions, but in use of newer methods (*e.g.*, FFT-based receivers, segmented estimation, 1-sec. carrier phase changes, etc.) while abandoning *habits* such as

- dismissal of partial fix data
- preoccupation with instantaneous position
- preference for location pseudomeasurements rather than the measurements themselves
- reliance on proprietary software and equipment "boxes"
- *r-f* interrogation/response sequences instead of squitters.

This paper was prompted by recent attention being given to aging satellites, but the points discussed herein can continue to be relevant in the future. Even with far greater coverage from GNSS, crises could emerge from severely stronger interference levels or other unforeseen events. Advance preparation for any such occurrence would avoid the waste, confusion, and blind alleys that generally arise with the sudden appearance of an emergency.

The industry can either adopt changes or continue to settle for performance levels at a minor fraction of the intrinsic capabilities available from our present and future systems.

REFERENCES

1. *Vulnerability Assessment of the Transportation Infrastructure Relying on the Global Positioning System*, VNTSC, August 29, 2001
2. *GPS World*, June 2009, pages 6 and 8
3. P. Nisner and R. Johannessen, "Ten million data points from TSO-approved GPS receivers: Results of analysis and applications to design and use in aviation," *IONJ*, Spring 2000, *pp*. 43-50.
4. Kalman, R.E., "A new approach to linear filtering and prediction problems," *ASME Journal of Basic Engr.*, March 1960, *pp*. 35-45.
5. Babu, K. ,*et. al.*, "Recommendations for a generic receiver interface," ION GPS-93.
6. Stoddart, L., "Open architecture, a new concept in vehicle location systems," ION GPS-94.
7. Marti, D., "GPS babel," *Linux Journal*, June 2002, *p*.22.
8. Farrell, J.L. and vanGraas, F., "That all-important interface," ION GPS-90.
9. Farrell, J.L., *GNSS Aided Navigation and Tracking*, Amer. Literary Press (NavtechGPS distr.), 2007.
10. Young, R.S and McGraw, G.A., "Fault detection and exclusion using normalized solution separation and residual monitoring methods," *IONJ,* Fall 2003. *pp* 151-169.
11. Farrell, J.L., "Robust design for GNSS integration," IONGNSS 2008
12. vanGraas, *et. al*. "Comparison of two approaches for GNSS receiver algorithms: batch processing and sequential processing considerations," ION GNSS-2005.
13 van Sickle, G.A., "GPS for military surveillance," *GPS World*, Nov. 1996.
14. Bayliss, E., et al, "Aircraft surveillance based on GPS position broadcasts from Mode S beacon transponders," ION GPS-94.
15. Farrell, J.L., McConkey, E.D. and Stephens, C.G., "Send measurements, not coordinates," *IONJ*, Fall 1999, *pp*. 203-215.
16 Farrell, J.L. and Farrell, M.L., "ADSB (2nd)-best foot forward?", *Air Traffic Control Journal*, Summer 2008.
17 Laurence J. Peter, *The Peter Prescription* , Amazon
18. Farrell, J.L. and vanGraas, F., "Integrity testing for GNSS sole means," ION Nat'l Tech Meeting, Jan. 1994.
19. RTCA Paper No. 455-93/SC159-463, Washington D.C.
20. R. Lilley, letter to RTCA-TMC, June 5, 1995.
21 Farrell, J.L. and vanGraas, F., "Comments on 'Summary of RTCA SC-159 Working Group Activities'," *ION J*, Winter 1997-98, *p*. 497.
22. Legal Issues Panel, ION-GPS 2000.
23 Farrell, J.L., "A quantitative measure of integrity error," IEEE PLANS, 1994.
24. Farrell, J.L. and van Graas, F., "Statistical validation for GPS integrity test," *ION Redbook V*, 1998, *pp* 89-100.

APPENDIX: VALIDATION REVISITED

Hard evidence of certification improperly bestowed, noted earlier, is clear from [3]. Failures from *TSO-approved* receivers outpaced allowable levels by *four orders of magnitude*. One-to-one correspondence of those results to everyday operation was not claimed but, by any reasoning, the conclusion is inescapable. Furthermore that was not the only example; spectacularly inadequate integrity capability of the first-ever certified GPS receiver is now widely known. This writer has no difficulty whatever believing that those early problems were corrected – that is not the intended focus of this Appendix. The point being scrutinized is *evaluation* – not for the purpose of criticizing but to point toward a far better way. After an unflinching look at test limitations, a key step toward solution will be described.

Critique of Current Practices

If today's validation criteria and methods produced stellar results, there would be no strong case to support change. Given the evidence, however, safety demands nothing less than a harsh and glaring spotlight cast upon today's test imperfections. That can pave the way for acceptance of major improvements. For those wishing to maintain the status quo, an initial note can dispel complacency. Existence of sophisticated methodology (RAIM) supported by accumulated experience – now over a period of decades – can lead to perceptions of a fail-safe system. Alas, any such complacency calls to mind the *Titanic*'s lifeboats, plus a statement from [17]:

When fail-safe systems fail, they fail by failing to fail safe.

Violation of containment boundaries – *if detected* – could lead to considerable delay and / or a lapse in operation. Undesirable as that is, it pales in comparison to an *undetected* violation. That raises the following concerns, reviewed here as succinctly as thoroughness allows.

- There is no standard integrity test; suppliers are allowed to devise their own validation methods. That may have been adequate for simpler systems of the past.
- Standardized blind testing (wherein those performing the test do not know the correct answers) was proposed and rejected by the collective will of the Fault Detection / Fault Isolation/Exclusion (FDI/FDE) Working Group for RTCA SC-159 (GPS Integrity). A paper [18] coauthored by co-chairmen of that Working Group advocated rigor in several areas of validation. Special attention focused on obvious failures produced the following common-sense prescription:

  "*Retest: ••• if the equipment being tested fails ••• , equipment must be modified to correct the problem before re-testing ••• .*"

- The following tract from [19] was highly instrumental in the rejection of the test plan just described:

  "*If a properly designed receiver fails the test, the manufacturer is required to modify or correct this receiver before retesting ••• This does not make sense: the receiver is, after all, designed properly, so what can the manufacturer `modify' or `correct' ?*" §

The self-evident flaw: a receiver <u>whose only outward sign is failure of a test</u> is assumed to be properly designed. Nevertheless, insufficient requirements were prescribed – and remain prescribed – for end-to-end (*r-f* in to final output) testing. Furthermore, even the software verification relies largely on pseudocode (for *integrity*; note the irony).

To ensure that the danger of a disastrous decision is fully confronted, some very inconvenient facts are highlighted. Letters written in response to those events included

- a letter from Ohio Univ. Avionics Engineering to RTCA-TMC advocating evaluation of test results "without commercial pressure affecting the outcome" [20] and
- a subsequent letter-to-the-editor of the ION Journal [21] also written by the FDI/FDE Working Group co-chairmen, raised the prospect of a technically unsophisticated but skilled marketing manufacturer.

Persistent doubts, expressed in[21], were finally vindicated by independent investigation [3]. In the extensive tests performed for that paper, failures were still unlikely – but obviously not unlikely *enough*. A shortfall in integrity clearly calls for thorough understanding of applicable quantitative requirements (*e.g.*, "how many nines" and why).

Shortly after appearance of [3], the following documented question was submitted to a Legal Issues Panel [22]:

"*Given the awareness of this situation ••• as well as the existence of documentation providing an example of misinterpreted certification test procedures, what are the liability implications for FAA, for the airlines, for the airframe manufacturers, and for the equipment suppliers in the event of an accident ?*"

The fact that no answer was recorded is also documented in those *ION-GPS2000 Proceedings*.

---

§ Interestingly, the source of the above quote subsequently made an observation that was 100% correct: Equipment that misses its requirement in a test trial, but by only a slight amount, should not carry the burden of a full penalty for failure. As applied to the context of that exchange (*i.e.*, integrity test), that same point had already been the basis of a documented investigation [23]. It is also the basis of further discussion for accuracy assessment, addressed later here.

The documented misinterpretation just mentioned refers to the first-ever certified receiver, failing spectacularly in multiple facets of integrity testing by another manufacturer as previously noted. Again it is readily acknowledged that correction of those early problems is quite credible, but one issue is inescapable: Historical proof of flightworthiness improperly bestowed did happen, which later became widely known. Instantly that calls into question the acceptance of proprietary rights for algorithms and tests.

Admittedly, simultaneous appearance of two or more of the various possible problems would still be unlikely. We can bend over backwards to acknowledge that – but there are no guarantees. The point being made is that "*unlikely*" *isn't good enough*. To illustrate this without undue complexity, an example is offered: consider a one-in-a-million chance for mishap. With 365 days in a year and over 3000 aircraft, each averaging four vulnerable flight phases (two takeoffs and two landings) per day, we could expect four mishaps per year from navigation. Unacceptable as that is, it only begins to address the overall scope of this issue.

To clarify the full impact of dire consequences that must be avoided at all costs, an unflinching look will now be exercised: Recall the meaning of *G* in *GPS*. Imagine hundreds of aircraft, carrying receivers validated by flawed integrity tests, all within the region sighting a flawed satellite whose position provides desirable geometry while some other satellites are not helpful to various users (due to outages, track loop interruption, multipath, blockage, sub-mask elevation, superfluous azimuth geometry, •••). There is no need to pursue the detailed results; except to say the stakes are so high that

- Failures to detect with risks on that scale need to be unlikely *in the extreme*; wildly improbable; *nowhere near* one in a million
- integration with external GNSS-independent data, with appropriate scaling from error statistics, could enable detections otherwise unnoticed.

The last item from among this topic carries the seeds of solution within it. First the problem: Efforts to obviate the limitations of GO/NO-GO integrity testing also failed to gain SC-159 FDI/FDE committee approval. As one result, consider a test with a maximum allowable number N of missed detections – irrespective of whether each may be a near-miss or a *blunder* – with the following hypothetical outcome from a large number of trial runs for two receivers (page 127 of [9]):

- RCVR #1 produces N missed detections, each occurring with errors exceeding allowable levels by orders of magnitude. Decision : *Accept*
- RCVR #2 produces $N + 1$ missed detections, each occurring with errors exceeding allowable levels only slightly. Decision : *Reject*

A Step Toward Solution – NO GO/NO-GO

The deficiency just noted was never corrected. Fortunately, remedial methods are available. Among those are

- probability scaling as noted in [18] to help substantiate extreme low probabilities
- Quantitative measure of error (sample mean and s.d.)
- abandonment of GO/NO-GO methods to set requirements

Binary (GO/NO-GO) choice criteria are applied in different places for very different reasons. Many in-flight decisions are "yes-or-no/this-or-that" rather than "how much." Plans for pilot training to change that are nowhere in sight. For test, however, the same rationale does not apply. The amount of error present in each test trial is known to the test equipment; a binary decision criterion replaces the range of values by a one-bit representation. A set of 5000 test trials with 0.001 allowable failure probability is described in [24]. If detections should have occurred in all 5000 trials and five were missed, then the maximum likelihood estimate for missed detection probability is 0.001 – but with only 50% confidence; the unknown actual missed detection probability is as likely to be above 0.001 as below it. For 99% confidence even only one missed detection occurrence would be excessive.

Intuition is not enough to yield high confidence, especially with limited numbers of test trials. Applying this reasoning to determination of unknown RMS error furnishes a clarification. Consider an effort to establish high confidence that a stipulated level for containment (*e.g.*, 0.3 nmi) will not be violated more than 0.1% of the time (probability cannot exceed 0.001). In view of he preceding paragraph a set of 1000 GO/NO-GO trials with no violations would be very inadequate for high confidence. Suppose, however, the error produced in every trial was less than twenty meters. Using those quantified results to authenticate high confidence would still require further formalized steps but, clearly, a major advance is available. A binary (yes-or-no; certify or don't certify) decision is needed at the *end of a test series* – but, emphatically, not within each individual test trial.

As with numerous other changes advocated in this paper, the requisite methods have a solid foundation and have been accessible for years (even decades).