Behavior and Social Issues, 15, 244-252 (2006).

RESPONSE AND COMMENT

MYTH AND REALITY IN WAKEFIELD'S ASSERTIONS REGARDING PAUL AND LENTZ (1977)

Gordon L. Paul[1]
University of Houston

I wish to thank the editor of *Behavior and Social Issues* for allowing me to respond to Professor Jerome C. Wakefield's extraordinary focus on the study reported by Paul and Lentz (1977) in his commentary on the Wyatt, Midkiff, and Wong articles in this issue (2006). Wakefield's pronouncement in his abstract and devotion of more than 28% of text to our book in a commentary on two articles—neither of which involve Paul or Lentz as authors—indicates an unusual interest in our report. This unusual interest is further indicated in the "*Myth and Reality*" heading under which most of his comments appear—a title I have borrowed as appropriate to this rebuttal.

I had listed nearly two manuscript pages of statements from Wakefield's commentary with which I agree. Unfortunately, because his commentary, itself, fails to reflect these points of shared "reality," space limitations require me to emphasize only the "myths" of his assertions, lest his faulty claims be accepted by the unfamiliar. Wakefield's overall critique is typified by a lack of theoretical understanding of his presumed theoretical enemy and its evidence base, selective perception to a degree suggesting harmful dysfunction, out-of-context quotes, and arguments based on factually incorrect allegations—clearly so in his assertions regarding the procedures and evidence in the Paul and Lentz (1977) monograph. Paraphrasing his own declarations regarding "behaviorists," to preserve his belief system, Wakefield seems in danger of turning claimed scientific critique into pseudoscience defended by *ad hoc* hypotheses and blatant falsehoods.

In the section preceding the one devoted to our report, Wakefield asserts, "There is a simple and quite scientific reason why token economies were set aside in favor of drugs; lack of generalization" (p. 211). He quotes Glynn (1990), "Maintenance and generalization of treatment gains are, of course, critical tests of the utility of treatment interventions" (p. 401) as if it were an indictment rather than the generally accepted premise of most clinical investigators (e.g., see Paul, 2006). Glynn's note that whether token economies were primarily prosthetic or therapeutic "has not been fully determined"

[1] Gordon L. Paul, Ph.D., is the Hugh Roy and Lillie Cranz Cullen Distinguished Professor of Psychology and Director, Clinical Research Program, Department of Psychology, University of Houston, Houston, TX, 77204-5022. E-mail: gpaul@uh.edu.

was followed by the assertion that this was a "sober assessment" and that "nothing has happened since 1990" to change it. He went on, "To underscore the reality of the generalizability problem, it is worth citing some of the *key empirical literature, without any pretense to a complete or balanced review*" (p. 212; emphasis added). The "key literature" included only early case reports and small developmental studies from the 1960s and 1970s, focused mostly on delusional speech—a cognitively-based problem that Paul and Lentz (1977) found, to obtain cross-situational generalization (with or without tokens), required longer courses of instruction than were in these reports. I will leave it to the paper's author to rebut Wakefield's claims in that section, but it seems clear that a constructed "straw person" was the focus there.

So that "things unsaid" are not misconstrued, before addressing the errors in Wakefield's assertions regarding our report, let me clarify my own position (see Paul, 2001). I am not an apologist for "behaviorists" and explicitly disavow "pigeonholing" in *any* "school." Because humans are electro-biochemical organisms and sentient social beings, a biopsychosocial approach makes sense for analyses of our behavior—an approach that includes possible genetic, biological, psychological, sociological, and environmental influences *and their interactions*. In the mental-health field, laboratory-derived principles of learning and performance along with physiological reactions to stress can account for the majority of problems that result in "symptom" labels for physically healthy individuals. Labeling problems as "symptoms" or diagnostic tags as "diseases," here, usually involves unproven theoretical constructions. People can be "prepared" to excel or breakdown by genetics, diseases, or environmental experiences that change their rate of response and the ultimate state achieved without nullifying the principles of influence in other spheres.

Further, a *utility criterion* is the best guiding heuristic in the clinical area. Because we nearly always deal with functioning after the point of development, our theoretical suppositions regarding how problems develop and how they are maintained must be hypotheses rather than facts. When more than one set of principles can explain a given phenomenon, the utility criterion directs us to invoke, not only the simplest set of principles *from those that could account for the data*, but also those that suggest ways of changing problem functioning as well. This results in hypotheses for ongoing testing to see if the formulation has predictive value (see Paul, 2001).

Wakefield introduces the specific section on our report as follows. "The classic Paul and Lentz (1977) study…has achieved a deserved iconic status as the most important study of token economy effects on schizophrenia….it is important to look carefully at what the Paul and Lentz study did and did not demonstrate" (p. 213). Two laudatory paragraphs ensue about the study and its findings, with extensive quotes from earlier reviews by Liberman (1980) and Glynn (1990). In Paul and Lentz, we anticipated Wakefield's next steps as well as his overall critique when we wrote, "…procedural descriptions, findings, and recommendations will predictably be quoted out of context and misinterpreted to bolster some emotionally laden argument" (1977, p. 443).

Wakefield slips in a truth and some half-truths before descending into misinformation and declarations based on factually incorrect allegations. Quotes and

clarifications are provided below to show how his selective reporting and typical failure to provide any appropriate contexts reveal his biases—even in the rare circumstance when his statements are not technically incorrect.

Following the laudatory paragraphs, Wakefield writes (p. 214), "However, despite intensive intervention over a long time in a controlled environment, the changes were not achieved easily and target behaviors were by no means eliminated. For example, after almost two-and-a half years, 'the average social-learning resident had improved to the point of demonstrating normal self-care and interpersonal skills *about half the time*' (p. 425; emphasis added). Nor were all spheres of behavior equally responsive; as Liberman (1980) notes, 'Even behaviorists will be disappointed to discover the failure of reinforced sampling-exposure procedures…(p. 369).'"

Explication of necessary contexts from Paul and Lentz (1977) that Wakefield excludes may help correct the connotations of his just listed statements and in his later critique. The *initial level and nature of patient disability,* as detailed in Chapter 12, provides an important context for any statement about rates of improvement. The patient groups were the most severely debilitated *ever* to be the subject of a systematic treatment study, averaging 17 years in the confines of mental hospitals—nearly two-thirds of their adult lives—and rejected for community placements just before participation. These were the individuals with whom mental institutions had unsuccessfully tried to cope, being chronically institutionalized even though *100% had received years of psychotropic drug treatment*. Among the patients randomly assigned to receive the Social-Learning Program (SLP), large proportions had received multiple series of convulsive shock "therapies" (50% electroconvulsive shock; 28.6% insulin shock), half were mute or incontinent, and many had spent day-after-day in canvas-restrained "hydrotherapy" tubs. Before the SLP began, for a full week, the *average* resident demonstrated bizarre functioning on 93% of hourly observations and *failed* to perform normal self-care and interpersonal skills, respectively, on 87% and 96% of opportunities.

Returning to Wakefield's statements (p. 214) with this added context, only the most naïve and clinically inexperienced commentator would expect that *any* treatment program could "easily achieve changes," "eliminate all target behaviors," or find "all behavioral spheres equally responsive" with such a severely debilitated group. Wakefield follows this bit of chicanery with an incomplete quote regarding improvement of the average resident to "normal self-care and interpersonal skills *about half the time* (emphasis added)." His emphasis neglects the fact that "about half the time" represents a sustained *increase of 377%* from initial levels of self-care and a remarkable *1200% increase* for interpersonal skills! Wakefield also cuts a clause, "*with the trend being toward continued improvements*" from our quoted sentence (Paul & Lentz, p. 425) and fails to mention that three residents had successfully qualified for discharge to independent living by this time. The final half-truth in Wakefield's paragraph above is the quote from Liberman (1980) regarding the failure of sampling-exposure procedures, which does reflect findings from the *first of two substudies* on this topic. Wakefield fails to mention that our second substudy (p. 247), testing improved procedures, was successful. A careful or balanced scientific critique? Hardly!

In the remainder of the section on the Paul and Lentz report, Wakefield's critique descends to a level of misinformation, declarations based on factually incorrect allegations, out-of-context quotes, and convoluted conclusions that are unequalled in any published "scientific" document that I have seen in more than 40 years in the field. With such extensive quotes from secondary sources, it is hard to tell whether Wakefield's misinformation comes from the misinterpretations of others or just from his own faulty scholarship. In any case, rebutting this misinformation on a line-by-line basis is precluded by space limitations (our original monograph, itself, is 528 double-column pages). Consequently, I will attempt to counter only Wakefield's most flagrant errors in the remaining pages here. I recommend that the interested reader examine the following selected publications and the references therein to help discern the actual facts involved (Mariotto, Paul, & Licht, 2002; Paul, 1987, 2000; Paul & Menditto, 1992; Paul, Stuve, & Cross, 1997; Paul, Stuve, & Menditto, 1997; Paul, Tobias, & Holly, 1972; Rhoades, 1981).

Wakefield's faulty assertions regarding supposed "methodological disasters" during follow-ups, incorrect information regarding deterioration of some residents immediately following release, questioning "why the benchmark for release was not lower," and attributing the high rate of community tenure to medication (see pp. 214-5) simply shows a failure to read or understand the facts in Chapters 35-43. The *only* methodological problem with follow-ups was that we were unable to carryout a planned comparison of *two different modes* of declining-contact aftercare based on social-learning principles (case vs. program consultation). This was because community personnel blended the two. Temporary declines in functioning for dischargees from inpatient programs (*only* those who experienced a reduction in social stimulation and an indiscriminant increase in prescribed drugs) were reversed once social-learning aftercare was begun as part of the planned generalization training to community environments and natural support systems. Community personnel, themselves, established the "benchmark" of functioning acceptable for community placements from all three inpatient groups; no "lower levels" were allowed!

Wakefield's declaration that the 97% success rate for dischargees remaining in the community "was based on the medication" reveals the most ideologically biased of all possible conclusions (p. 215). Consider the facts. 100% of all participants had been on psychotropic drugs for years—yet they remained chronically institutionalized. A triple-blind withdrawal study over the first few months following introduction of the SLP and Milieu Programs (Paul, et. al., 1972) found the only difference between equated drug-continuation & drug-withdrawal subgroups in both psychosocial programs was that those continuing on drugs *took longer* to respond to psychosocial treatments. The highest rates of improvement occurred during periods with the fewest patients remaining on drugs and the greatest deterioration occurred *following* the greatest number of patients being placed on drugs. Holding drugs constant during the last 6 months at fewer than 11% for the SLP and 18% for the Milieu Program was associated with renewed improvements. The Hospital group had the lowest release rate, with 100% of patients receiving psychotropic drugs (51% at high dosages). Drug status *never* predicted positive change in 6 years of

data collection with these patients. How indiscriminant prescription of drugs to about 80% of community placements would suddenly account for success seems to require magical thinking!

Wakefield's (p. 215) misquote of Paul and Lentz (p. 410) actually refers to our inability to draw further conclusions about the two modes of declining contact psychosocial aftercare based on social-learning principles rather than having *anything* to do with generalizability. On the same page we write, "…the major impact appears to have come from the introduction of active psychosocial aftercare consultation, no matter what the mode of delivery" (p. 410). The success of our declining contact aftercare approach in maintaining dischargees from all three groups with rehospitalization rates of less than 3% over periods from 18 months to more than 5 years is strong evidence for it's influence. This is especially so when compared to rehospitalization rates of about 59% within 12 months for such placements who received the usual drugs and aftercare services during the same period elsewhere in the state (Rhoades, 1981, p. 3). *Such generalization training to community environments and natural support systems does appear to be the key to community tenure, no matter how patients achieved the level of functioning that qualifies them for discharge.*

Later I will provide evidence contrary to Wakefield's completely unsupported declaration that the Paul and Lentz study "fails to demonstrate (or even to test) generalizability" (p. 215); however, his misrepresentation of the timing and procedures for dealing with assaultive behavior (p. 216) is so replete with partial quotes and misquotes from secondary sources and factually incorrect statements (e.g., "Draconian punishment") that I must clarify that record. It should be clear that the aversive aspects of the SLP play an important, but limited role in the total program, which is a complex interactive package that is overwhelmingly devoted to positive, constructive actions (see Chapters 42 & 43). Measures for dealing with assaults may seem extreme to those without direct experience with patients as severely disturbed as those in the original comparative study. "For example, assaults ranged from the ordinary bodily blow to biting a finger off, breaking a leg, and literally hurling a staff member over a counter a distance of some 10 feet" (p. 493).

Following the above incidents, with understaffing often resulting in only two female staff on each 28-bed ward during evening- and night-shifts, an increase from 2-hours as a consequence for assaults to 72-hours for Time-Out (TO) in seclusion ("expulsion" in the Milieu Program) was introduced midway through the second six-months of operation. Contrary to Wakefield's report, the 72-hour TO period for assaults was in effect for only three months, and then reduced to 48 hours. As a result of a statewide mandate, it was, again, reduced to 2-hours just before the 5th six-monthly anniversary (with both programs showing high rates of improvement). The latter reduction to 2-hours remained in effect for more than 18 months, until return to baseline procedures were instituted for four weeks. Contrary to Wakefield, programs were reintroduced for the last six-months with TO for assaults being from the time of an incident until 5:30 AM the next day (i.e., up to a 23.5-hour maximum). Wakefield fails to mention that, when staffing is adequate, overcorrection/restitution procedures are 100% effective in eliminating assaults that

TO/token-fine procedures fail to control (p. 462). Of course, all of these procedures received review, approval, and oversight from committees at the local facility, state, and federal levels. High-dose drugs were used only for the most assaultive residents before institution of the longer periods of TO.

Even with the above intrusive contingencies, the SLP is by far the most positive of any ward-wide program for severely disabled clientele, with a positive-to-negative ratio of more than 23:1 for material goods and services (3:1 in other programs). By actual measurement, negative staff-resident interactions with these most-disturbed people occurred at a rate of 7.5% for the SLP, 12.3% for the Milieu Program, and 20.9% for Hospital comparisons. Far from Wakefield's "Draconian punishment," the defined and assessed performance criteria and contingency rules ensure the least restrictive means necessary within the program itself—including freedom from unnecessary psychotropic drugs. In fact, "the SLP as recommended provides the least intrusive treatment for each resident, nearly automatically" (Paul & Lentz, p. 463).

Wakefield completely misrepresents the degree, nature, and amount of loss for residents in the SLP following the statewide administrative reduction in time-out for assaults to 2 hours (216). By selectively reporting only the weakest measurement of all those involved, he claims the SLP showed "regression after years of treatment even as treatment continued"—completely disregarding three pages of analyses (pp. 377-379) that show how nearly everything he asserts as facts regarding absolute levels of functioning and comparative change are *actually artifacts.* The artifacts were due to clinical staff raters shifting their level of patient's ratings up or down on the basis of their own temporary emotional states. In an unbelievably prejudiced rejection of evidence, Wakefield (p. 217) dismisses as "a bit more positive in absolute terms in later assessment periods" the findings from the objective assessments of functioning that have been consistently documented to retain absolute levels of measurement over places and times (see Paul, 1987; Mariotto et. al., 2002).

Most of Wakefield's above claims, in fact, are undercut by the objective data from the Time-Sample Behavioral Checklist (TSBC), gathered on stratified-hourly observations by trained independent observers, and from the Clinical Frequencies Recording System (CFRS), completed moment-to-moment by clinical staff (Chapters 24, 27, 42, 43). To be sure, the arbitrary reduction of the length of time-out and expulsion from 48 to 2 hours had a catastrophic impact on the Milieu Program and halted the trend of greater continuing improvements for the majority of residents in the SLP. However, contrary to Wakefield's assertions, this "natural experiment" actually provided evidence *for* the generalizability of previous gains to different environments for SLP residents.

The group of SLP residents continued to show significant rates of improvement from their original levels of functioning throughout the 18-month period without effective consequences for assaults. Previous gains over initial levels were maintained in every area of objectively assessed functioning, *except* assaultive behavior. Because the uncontrolled increases in assaultiveness obviously resulted in a less therapeutic environment for everyone, maintenance of previous gains in a less hospitable environment is, arguably, evidence of generalizability.

Direct evidence of the generalizability of previous gains to a different environment was obtained from ongoing objective assessments of all areas of functioning during the four weeks in which return-to-baseline procedures removed all components of the SLP (tokens, programmatic material and social reinforcement, TO, etc.). This was done explicitly to test generalized improvements from the "old program" before reinstituting the new programmatic procedures, during the last six months of operation (Chapter 28, 29, & 31). Except that drugs were held constant at 11% during the 2nd one, assaults were handled by traditional hospital means during both the 1st and 2nd baselines—restraints, tepid baths, physical separation, and instruction. Staff were told to be pleasant, but to emulate the "aid culture" during both baseline conditions. The "82% improved" figure for SLP residents during return-to-baseline, following 18 months of operation without effective means of dealing with assaults, is *strong* evidence of generalization.

As a final challenge to Wakefield's unsupported declaration that "the Paul and Lentz study fails to demonstrate (or even to test) generalizability" (p. 215), I note the following. The SLP manuals are *based on years of clinical/scientific work by hundreds of researchers* so that the development of proficiencies essential to successful community living follows successive approximations to "normal" functioning at every step. Generalization training across people, times, and settings as well as to less structured environments is built into the program, with systematic fading of artificial training procedures as natural support systems become functional. Hourly observations document that staff performed these procedures with less than one-half of 1% error in all interactions, and objective ongoing data document generalization within the program. Following declining-contact aftercare, objective assessments during the last week before discharge predicts in-community ratings of functioning at 6, 12, ans 18 month follow-ups, with correlations in the ±.60s and ±.70s—evidence of generalization unprecedented in the institutional literature.

Granting that Wakefield is completely wrong in his claims about the absence of evidence for generalization of improvements from the SLP, it should be clear that a "token economy" is not a treatment in and of itself, but simply an organizational delivery system. The SLP reported in Paul and Lentz is a comprehensive program that specifically builds in delay of gratification and generalization training to less structured environments and natural support systems. Token-economy programs that do not articulate all these necessary details, including ongoing assessment procedures, "may be no more related to other token-economy programs than the action of heroin is related to that of penicillin, even though both are administered by injection" (p. 434).

Elsewhere (Paul, 2000, p.7), I highlighted the current status of the evidence for a fully implemented SLP with integrated declining-contact aftercare as follows.

- "100% effective in improving functioning for the most severely disabled people with mentally ill diagnoses and lengthy hospital stays—their clear 'treatment of choice.'
- Highest rates of successful discharge (>90%).
- Lowest rates of recidivism or rehospitalization (<3%).

- Empowers clients through self-determination and competency training.
- Ensures the least-restrictive interventions for each individual.
- A promising approach for less severely disabled clientele with shorter hospital stays.
- Self-improving by ongoing discovery/incorporation of emergent evidence-based procedures.
- Applicable in public and private hospitals, mental health centers, and community facilities.
- Requires no greater numbers or levels of staff than are present in most existing operations.
- Cost-efficiency that is better than three times that of alternative approaches."

The above evidence is part of the reason the SLP and our assessment technologies were included in the APA/CAPP (2005) "best-practices" for people with serious mental illness. Using words from elsewhere in his own commentary (p. 210), Wakefield's attempts to discredit this evidence appears as "a bewildering assessment" when "we are talking about life-destroying and family-destroying" conditions in which such evidence "in any other…domain…would be cause for joy, and would be considered justification for recommending universal administration." Wakefield's concluding characterization of the targeted papers (p. 219) appears to be an accurate description of his own commentary, as he seems "to be deploying all the techniques of which pseudoscientists have been justly accused in the past." His selective emphases, "whatever their rationale, have the appearance of the zigzagging of a pseudoscientific defense of an ideology, not a balanced scientific assessment of the evidence."

REFERENCES

APA/CAPP Task Force on Serious Mental Illness and Severe Emotional Disturbance (2005). *Training grid outlining best practices for recovery and improved outcomes for people with serious mental illness, Version 1.3*. Retrieved July 16, 2006 from http://www.apa.org/practice/grid.html

Glynn, S. M. (1990). Token economy approaches for psychiatric patients: Progress and pitfalls over 25 years. *Behavior Modification*, *14*, 383-407.

Liberman, R. P. (1980). A review of Paul and Lentz's psychological treatment for chronic mental patients: Milieu versus social-learning programs. *Journal of Applied Behavior Analysis*, *13*, 367-371.

Mariotto, M. J., Paul, G. L., & Licht, M. H. (2002). Assessment in inpatient and residential settings. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches (2nd ed.)* (pp. 466-490). New York: Oxford University Press.

Paul, G.L. (Ed.). (1987). *Observational assessment instrumentation for service and research–The Time-Sample Behavioral Checklist: Assessment in residential treatment settings, Part 2*. Champaign, IL: Research Press

Paul, G.L. (2000). Evidence-based practices in inpatient and residential facilities. *The Clinical Psychologist, 53*, 3-11.

Paul, G.L. (2001). The active unconscious, symptom substitution, & other things that went 'bump' in the night. In W.T. O'Donohue, D.A. Henderson, S.C. Hayes, J.E. Fisher, & L.J. Hayes (Eds.), *The history of the behavioral therapies: Founder's personal histories* (pp. 295-336). Reno, NV: Context Press.

Paul, G.L. (2006). Psychotherapy outcome can be studied scientifically. In S.O. Lilienfeld & W.T. O'Donohue (Eds.), *The great ideas of clinical science: 17 concepts every mental health researcher and practitioner should understand* (pp. 117-145). New York: Brunner-Routledge.

Paul, G.L, & Lentz, R.J. (1977). *Psychosocial treatment of chronic mental patients: Milieu versus social-learning programs*. Cambridge, MA: Harvard University Press.

Paul, G. L., & Menditto, A. A. (1992). Effectiveness of inpatient treatment programs for mentally ill adults in public psychiatric facilities. *Applied & Preventive Psychology: Current Scientific Perspectives*, *1*, 41-63.

Paul, G. L., Stuve, P., & Cross, J. V. (1997). Real-world inpatient programs: Shedding some light. *Applied & Preventive Psychology: Current Scientific Perspectives*, *6*, 195-206.

Paul, G. L., Stuve, P., & Menditto, A. A. (1997). Social-learning program (with token economy) for adult psychiatric inpatients. *The Clinical Psychologist*, *50*, 14-17.

Paul, G.L., Tobias, L.L., & Holly, B.L. (1972). Maintenance psychotropic drugs in the presence of active treatment programs: A "triple-blind" withdrawal study with long-term mental patients. *Archives of General Psychiatry, 27*, 106-115.

Rhoades, L. J. (1981). *Treating and assessing the chronically mentally ill: The pioneering research of Gordon L. Paul.* DHHS Publication No. (ADM) 81-1100. Rockville, MD: Science Reports Branch, Division of Scientific and Public Information, National Institute of Mental Health.

Wakefield, J. C. (2006). Is behaviorism becoming a pseudo-science?: Power versus scientific rationality in the eclipse of token economies by biological psychiatry in the treatment of schizophrenia. *Behavior and Social Issues, 15,* 202-221.