First published in New Ideas in Psychol. 1995, Vol. 13, No. 3, pp. 237-245. A response to Kukla (1995).

Towards More Effective Metatheory: A Response to Kukla

There is much to admire in Kukla's essay, especially his advocacy of taking metatheory (disciplined study of theories) seriously, the earnestness of his effort to articulate a technical framework for conducting such inquiry, and the instructive fashion in which this has accomplished so little.

I say this not to be cutely malicious, but because what I shall claim to be deficient achievement here is instructive if viewed from a concept-engineering perspective. We can learn much from postmortems on unsuccessful endeavors, and scientific psychology's now-primitive metatheoretic proficiency still has enormous growth potential that I, for one, would dearly love to see more fully realized.

There are at least three levels-interlaced, to be sure-on which one can pursue scientific metatheory: (1) At the substantive level, we can critique extant theories (in one sense or another of this protean notion[1]) that have been seriously proposed for the sake of what they say. This substantive target of appraisal can be the critic's own problematic explanation for certain phenomena, or it might comprise proposals of others currently competing for approval. But the goal of substance-focused metatheory is to evaluate and perhaps improve upon ideational material in a science's active workspace. The Chomsky vs. Putnam debate, Spence on transpositional learning, and Fodor's token-identity proposal are illustrations proferred by Kukla of metatheory at this level. (2) Methodological metatheory seeks to develop theory-appraisal concepts and procedures that are relevant to, and increase our prowess at, substantive metatheoretic inquiries over a broad range of applications. Kukla's expansions of the themes invoked by his section headings illustrate this level of metatheory. (3) Finally, on the foundations level, large operational questions remain about the nature of human thought and its relation to the external realities that test its worth. By "operational" here I mean to wave off impersonal studies of cognitive processes in college undergraduates, or maxims of approved thinking in philosophy textbooks, and focus on efforts to learn more about what we really do, and how we might do it even better, when as professional epistemic engineers (which is what most scientists and philosophers are), we are productive at top of our form.[2] I shall argue that Kukla's present metatheoretic

[1] Conceptual systems have many distinguishable aspects that we commonly construe as "theoretical" (see Rozeboom, 1970, pp. 59-73 for an inventory considerably more diverse than Kukla recognizes here) but all belong within the purview of metatheory.

[2] Note the 1st-person pronouns here. Level 3 metatheory embodies existential epistemic con-

concerns at Level 2, while by no means trivial, have for the most part only minor relevance for the work that so massively and importantly needs doing at Level 1 because they ignore the most salient problems at Level 3.

Kukla builds his discussion of theory comparisons upon a probability measure $p(\ )$ whose arguments in this application are theories or totalities of their empirical consequences. That is (though Kukla does not say so explicitly), his $T_1, T_2, \ldots$ here are conjunctions of sets of declarative sentences that constitute specific theories or their empirical consequences. Kukla acknowledges that $p(\ )$ over theories is an idealization which some philosophers disavow. But he seems not to recognize that this idealization (which he does take seriously after promising not to) brutally suppresses the leading problems of theoretical praxis in science. I had thought to commence this protest by pointing out that the probability of propositions (subjective credibility) differs profoundly in logical kind from the probability of attributes (objective statistical probability), that its classical axioms mirroring those of statistical probability are at best normative ideals that real-world belief systems can never be expected to approximate even roughly, that the propositions it presumes as arguments are likewise perfections far distanced from the contents of ordinary or even extraordinary human thought, and so on ... ending in the posit that even if classical credibility theory correctly axiomatizes the structure of an ideally coherent belief system (which may well be true), the foremost goal of an operational theory of rational belief management is learning how the dynamics of our perceivings/conjecturings/intuitings/inferrings/etc. can best be self-controlled to reduce our estrangement not merely from ideal coherence but even more importantly from other perfections without some decent approach to which coherence is vacuous. However, leisurely unfolding of such meta-metatheoretic abstractions is impractical here. Let's jump to the bottom line.

The bottom line is that Kukla's present theory comparisons presume that its T_1, T_2, etc. are fully in hand as published statements with articulately coherent logical forms and unambiguously clear meanings. As attested by his remarks under Amplifiation, Kukla is by no means unaware that theories in practice may prove wanting by these standards. But he disregards how profoundly pervasive are such defects. All the Level 1 theoretical materials with which I have worked in psychology, and a lot in philosophy, comprise just fragmentary simulacra of sentences ready for credibility appraisals. To warrant $p(\ )$ ratings, these first need their ellipses to be unpacked, their vagaries alleviated, the scopes of their generalities delimited, their ambiguities of reference resolved, and their non-sequiturs mined for hidden presumptions. My early work on behavior theory documents many examples of such need for conception therapy, some utterly fundamental for

cerns, not academically distanced studies in the history or sociology of science. When I spoke just now of empirical cognitive studies as "impersonal' I was reaching for an adjective that communicates disengagement, as in "Isn't it clever what those monkeys are doing!"

progress on the topic at issue. (See Rozeboom, 1958, on prejudiced description of conditioning phenomena; Rozeboom, 1960, 1961, on the corruptive inarticulateness of traditional behaviorist stimulus/response locutions; Rozeboom, 1965, on misdirected construals of memory; Rozeboom, 1970, pp. 103-108, 130-136 on the hoaxiness of most S-R mediation models.) And although the easy explanation why none of this work ever received any attention by other behaviorists or their critics is that behavior theory was by then in the last twitch of its death throes, that does not account for our discipline's equally profound neglect of MacCorquodale & Meehl's 1954 gem of Level 1 metatheoretical analysis a decade earlier when behavior theory was still ascendant. A bleaker conclusion is that few research psychologists have developed either the analytic skills required to explicate a living theory's logical character or the motivation to acquire them. (I have urged the importance of breaking out of this poverty with youthful fervor and some eloquence in Rozeboom, 1961, and again, with less eloquence but more matured disillusion, in Rozeboom, 1977. The latter also recommends some self-help exercises which could expedite that breakout.)

Many of Kukla's proffered categories of theory comparison-amplification, internal consistency, intertheoretic entailment or inconsistency, etc.-address logical relations to which a conceptually advanced discipline's theory development should indeed be attentive. And if a researcher thinks that his favorite position statement on some topic is related in Kuklean fashion to certain competing proposals by others, this may well invigorate interchanges wherein one or more parties to the discourse instructively reshape their positions, perhaps even promoting conciliations, deepened understandings, and fresh ideas. (Don't hold your breath, but it does happen occasionally.) What this does *not* bring about, however, is altering of credences ascribed to these contending theses while leaving our verbalizations and understandings thereof just as they stood at debate's outset. Nor should that occur, for what would be gained by it? Imagine that during our future interstellar adventures, we have stumbled upon a God-like information-processing device, Theus, which receives sets of sentences and sometimes (though not always, since Theus refuses to bite on paradoxes) returns epistemic evaluations of the propositions we understand these to express. And we have also become convinced (never mind how) that Theus's judgments are never wrong. Now suppose that you and I agree that our respectively favored theories T_1 and T_2 appear to be in some disagreement about their common topic. If we present $T_1 \cdot T_2$ to Theus, would we really be much enlightened if Theus' response is "T_1 is entirely correct apart from a few minor blemishes, but T_2 is seriously in error"? Sure, we now both know that T_1 is right and T_2 is wrong; but neither of us (or at least not me and I hope not you) would take much epistemic comfort from that unless we can get some sense of how it is that T_1 is belief-worthy while T_2 has gone astray. Alternatively, if Theus disables our impression that T_1 and T_2 are incompatible by advising us that T_2

is in fact a misleadingly phrased special case of T_1, we would—or should—remain metatheoretically distraught until we manage to see, firsthand, how T_1 subsumes T_2.

In short, when appraising verbalized theory simulacra for their Kukla-connections, just about the smallest metatheoretic benefit to expect from the inquiry is some closer proximity of the credences we assign these word strings to the coherence of a classically ideal belief system. The big prize, if we are astute, will be enhancement of our thinking on this topic by expanding its ellipses, alleviating its vagueness, delimiting the scopes of its generalities, identifying the presumptions impelling its non seqiturs, and perfecting its conceptions in still other ways not mentioned above. This is not at all to suggest that we become indifferent to credibility. Just the opposite: by refining muddled notions into theses with full-bodied prepositional content, we can move beyond the mouthing of slogans to meaningful convictions or knowledgeable suspensions of judgment as deepened understanding reveals the evidence to warrant.

Perhaps I am being unfair to Kukla here. Since he has not expressly declared that helping scientists to theorize more effectively is one of his aims, possibly his intent is merely to discuss idealized theory relations without foreground concern for how those might be realized in practice. But I choose to put a how-to-do-it spin on his account, because that's where action is most needed.

My argument that Kukla's theory relations are generally not the salient focus for Level 1 metatheory has so far been abstractly bloodless. Let me call upon Kukla's own examples to show how their Kuklean outcomes were largely opportunities wasted.

Spence on transposition of learning

Kukla is right to see confirmational merit in Level 1 metatheoretic argument that phenomena previously thought to refute a certain theory are actually compatible with it. But the main lesson to take from his proffered example in Spence's theory of transposition in conditioned responding is rather less cheerful than that. The transposition phenomenon, which Kukla has nicely summarized, had seemed incompatible with classical S-R theory's principle of stimulus generalization, namely, that conditioning a response R, or inhibition thereof, to a stimulus S in some degree of strength d also generalizes R-doing, or R-suppressing, to all other stimuli S' in lesser strengths d' that diminish from d with decreasing similarity of S' to S. Spence's argument, that suitable combinations of specially shaped gradients of generalized evocation/inhibition can yield transposition, is cold comfort when the posited shape of those gradients is so implausible. But far worse, Spence's model was fundamentally inchoate. The response R_2 presumably conditioned to stimuli S_2 here is $\langle \mathit{approach}\, S_2 \rangle$. Hence under the received view of stimuli and responses, what generalizes to S_3 when R_2 is conditioned to S_2 is $\langle go\, to\, S_2\rangle$, not $\langle go\, to\, S_3\rangle$. So discrimination training just on S_2 (positive) paired with S_1 (negative) should never attach any $\langle go\, to\, S_3\rangle$ strength to stimulus S_3 at all, whence on test trials pairing S_2 with S_3 the subject's directly conditioned tendency to approach S_2 in response to S_2 should be intensified by S_3's generalized elicitation of this very same approach. And S_1-avoidance should not affect test-trial performance at all. Before behavior theory can assimilate transposition, it needs conceptions of stimuli, responses, and their evocative couplings under which conditioning of $\langle go\, to\, S_i\rangle$ to stimulus S_i can generalize the adaptively different response $\langle go\, to\, S_j\rangle$ to a novel stimulus S_j. Nothing in generic behavior theory nor even in the narrower S-R outlook save muddy thinking stood athwart that conceptual enrichment (see Rozeboom, 1960, 1961, p. 480f, 1974, p. 234f.; and its attainment would have advanced behavior theory to the threshold of structural complexities in the central mediation of overt behavior that not even modern cognitive psychology has yet adequately recognized. The value of Spence's transposition model was not in its tiny credibility support for then-standard S-R notions but in the explosion of behavior-theoretic concept analysis and foundation probing it should have ignited. That this explosion was a dud can be forgiven its era. But what are we doing to discourage such metatheoretic misfires from recurring today?

Token-identity theory

Modern American analytic philosophy has many strengths, especially its naturalistic outlook and its proficiency in sketching philosophical issues of considerable sophistication in bold strokes using simple materials from everyday language. But the flip side of its style is a proclivity to glibness that far too often treats an analysis as complete when it has scarcely entered the deep water. Fodor's distinguishing between "type" and "token" theories of mental/physical identities to separate materialism from reductionism is a good case in point.[3] Not merely has token-identity theory *not* "established the logical independence of the two doctrines" (even Fodor professed only a partial separation), it has not even shown the alleged distinction to have any philosophic or scientific significance albeit one could scarcely find a topic in whose vicinity more of philosophy's most basic issues come to simultaneous boil.

The Level 1 question here is when, if ever, an object x's having some mental property M is the same as some object y's having physical property P. Proposing answers to that is fatuous, however, until we have done some Level 3 foundation

[3] Kukla correctly notes that Fodor's distinction has precedents, most importantly in Davidson's "anomalous monism." But to my knowledge, Fodor was first to apply the type/token labels from Linguistics to it.

work on what such identity hypotheses might mean, and what maneuvers can lead to enlightened conclusions about them, even when M is not stipulated to be mental or P physical. Flipping quickly over the early pages of this inquiry, we find ourselves asking metalinguistically how different it is possible for two sentences '$\phi(a)$' and '$\psi(b)$' of an ideal language to be, in which 'ϕ' and 'ϕ' are predicates (open sentences) and 'a' and 'b' are nominals (nouns or noun phrases), and still both signify the same objective state of affairs. But to proceed from there we need to posit an ontology of what there is that an ideal language can represent, and a semantics of how its representations are accomplished. It should not surprise you to hear that this is still a matter of some dispute even for an ideal language, not to mention our imperfect real-world approximations thereto. But there is general agreement that ideally successful nominals designate objects (individuals). Beyond that, the prevailing opinion (or at least one that is common and, with some qualifications, shared by me) is that ideally successful predicates delimit classes of objects by designating their properties (attributes, features, characteristics), while an ideally successful subject/predicate sentence '$\phi(a)$' represents the fact (state of affairs) that a has property ϕ or—what some consider equivalent—the *event* of a's-having-ϕ which is what this sentence's gerundization ("a's-ϕing") designates. (Actually, there is reason to doubt the universality of this thesis; but until we develop some hard theory of its limits, we pretty well have to pretend we don't need them.) We all(?) agree that predicates have the same referent if they are synonymous, and only if they are co-extensive; but we still have no tightly reasoned doctrine on when or how two non-synonymous predicates designate the same property.[4] Even so, that need not deter us from assuming "type"-identity $\phi = \psi$ (the object-language extrusion of positing that non-synonymous 'ϕ' and 'ψ' are co-referential) as a premise of philosophical argument. Thus in particular, it is relatively unproblematic to stipulate that mental events consisting of individuals having mental property M are "type-identical" with physical events consisting of those same individuals having physical property P just in case $M = P$.

But what is "token identity"? For Fodor in this context, "tokens" were events constituted by some individual having some property, so identity of tokens has a different domain than does identity of types. But can a's-having-P be identical with b's-having-Q without requiring $a = b$ and $P = Q$? Fodor (1975, p. 13)

[4]For example, how can we decide whether, over the domain of geometric polygons, Triangularity is identical with Trilaterality? And how should we explicate our intuition that *Taller-than* and *Shorter-than* both refer in some fashion to the same relation even though 'a is taller than b' contradicts 'a is shorter than b'? In principle, developing theories of property identity should be reasonably straightforward once we have conceived a suitably rich repertoire of the meta-attributes that differentiate properties; however, that inventory is still largely bare. Moreover, causal/becausal relations must figure prominently in this still-untold story; and few philosophers seem to appreciate how empoverished our understanding of causality and non-causal explanatory dependencies still remains.

thought so: "Token physicalism does not entail type physicalism, if only because the contingent identity of a pair of events presumably does not guarantee the identity of the properties whose instantiation constitutes the events." But he never hinted at an argument for this presumption, and quickly moved on to an issue far more central to his concerns, namely, the prospect of mental predicates being in principle co-referential with physical ones. Perhaps he felt no need to support his "... does not guarantee ..." presumption because it seemed that a recent flurry of papers on the individuation of events (excellent work; no glibness there, or at least not much) had already settled this. But of course that had done no such thing; for foundations adequate to resolve the matter (like most others in ontology) have not yet evolved. What this work had mainly made plausible is that an event comprising x's having some highly determinate property P is frequently, perhaps in practice always, referred to by gerundized sentences whose predicates make explicit only a fragment of P. Thus if we know that John spoke harshly to Marsha last night, it is not unreasonable to understand "John's speaking harshly last night" and "John's speaking to Marsha last night" as both referring to John's speaking harshly to Marsha last night. But neither is it unreasonable to contend as well that "John's speaking harshly last night" refers in contexts intolerant of ellipses (say in testimony at a trial, with short-range demonstrative "last night" replaced by something like "on the occasion at issue") to a more abstract event, John's speaking harshly last night, embedded in the other. (Were space to permit, I'd gladly explain why this is not merely reasonable but perhaps mandatory for an ontology we can live by.) Moreover, even if event descriptors are allowed to omit details of their target event's character when context disambiguates their reference, this does not preclude that if an event e is *fully constituted* by x's having P and also by x's having M, then necessarily $P = M$. But if the latter is true, then only a short step remains (after *ism* explication I shall forego here) to conclude that materialism entails reductionism, contrary to what Kukla claims has been established by token-identity theorists.

The point of this breakneck review is my objection not to Fodor's being presumptuous in 1975—his original sundering of materialism from reductionism was scarcely more than a throwaway conceit—but to the supposition that seems henceforth to have prevailed among cognitive scientists and philosophers of mind, taking Fodor's word for it with no serious argument or analysis, that token-identity sans type-identity frees materialism (cheers) from the stigma of (shudder) reductionism. Why is this objectionable? Mainly because swathing a delicate research topic in secondhand slogans, like spectators trampling the evidence at a crime scene, degrades the prospect of progress therein. But also, it seems irrationally wimpish to fear that the autonomy of psychological science is jeopardized by type-identity of mental events with supervenient physical ones that are not now conceivable, nor ever will be, in the language of any science whose subject matter has traditionally been labelled "physical." Since Identity is symmetric, reducing mind to matter can just as well be viewed as reduction of matter to mind, especially if we are prepared to admit—and why not?—the prospect of mentality in grades far more primitive (in particular, at lower levels of supervenience) than those we introspect upon or infer from commonsense intentionality talk. Let's hear it for pan-psychic reductionism. Even so, feckless token-identity theorizing has probably done no real harm. The serious work has been underway next door on supervenience relations among properties. This is requisite to further progress on property identity and individuation of events, and appears to be in good hands, especially those of Kim (e.g. Kim (1993)). But we should look upon recent token-identity enthusiasms not as metatheoretic triumph, as Kukla seems to want it, but as a high-level philosophic embarrassment.

Chomsky vs Putnam on the origins of language

Since Chomsky's innateness-of-language thesis (IH) is so murky and Putnam's argument, that common-origin hypothesis CO is implied by IH, is so much shakier than Kukla acknowledges, I had initially thought to list this debate as yet another example of inept Ist-level metatheory. But review of its full text shows it to support my case from a happier direction. Putnam was not attempting to appraise IH's $p(\ )$-rating, else he would not have insisted at outset that "the I.H. seems to me to be *essentially* and *irreparably* vague" (Putnam, 1980, p. 242; his italics). Throughout his commentary, Putnam sought to prod Chomsky into taking the marbles out of his mouth. Impeccable deduction of CO from IH was not required by his game plan; what mattered was that Chomsky hated CO and thus might be goaded to clarify IH if even a soft argument could stigmatize that with an apparent commitment to CO. So just as I have been urging, Putnam was putting first things first: before sweating a proclamation's possible truth, try to decipher what its proponents may be trying to say.

Epilog: The evil of theoretic holism

There remains one other Level 2 metatheoretic issue implicitly raised by Kukla that has enormous import for scientific practice. This is the monstrosity of appraising theories as wholes.

Present commentary has so far emphasized the unwisdom of addressing scientific theories with premature concern for their credibility. But the holistic tenor of Kukla's theory appraisals would merit demurrer even were his Ts semantically ideal. Since he himself moves beyond that in a section I shall commend, I would forego this particular criticism were it not that theoretic holism does so much damage in psychological research that, like spruce budworm, kudzu, and politically correct ideologies, it needs vigorous opposition wherever it erupts. By "holism" here, I mean the attitude that a 1st-level theory is best treated as a single sentence with no internal structure beyond whatever minimal parsing may need recognition just long enough to establish some testable prediction or entailment/contradiction relation to other holistically treated theories. In particular, holistic theory adjudication sees changing a theory's credibility by impact of data or Kuklean contact with other theories as an operation performed on the theory in its macro-entirety rather than as an epiphenomenon arising from differential adjustments in the plausibilities of its micro-constituents. Students are instructed to think that way by the Popperian hypothetico-deductive account of theory evolution, which is usually the only metatheory on scientific inference to which they are exposed. And worse—because no aspirant psychologist can escape this brutalization—holistic hypothetico-deductive reasoning is hard-wired into the Null-Hypothesis Significance Test on which training in research methodology is almost always centered.[5]

Rozeboom (1970, pp. 91-103) develops the case against theoretic holism in considerable detail that need not be reviewed here. It suffices to suggest that Kukla's theory comparisons will be considerably more fruitful in 1st-level applications if they are focused upon these theories' individuated constituents than on the macro-theories as wholes. Gratifyingly, Kukla too recommends this finer grain of analysis when, in his discussion of Simplification, he encourages efforts to discern within a theory its minimalist essentials for explaining the data it aims to comprehend. Unlike Kukla, I see no need to raise a theory's credibility by expunging it of components that can be eliminated without loss of empirical import so long as we remain clear that these attachments have no evidential support and seem to be only cosmetic; while on the other hand, I urge that among a theory's empirically portentful components it is important to distinguish those that are supported or challenged by data already in hand from those whose adjudicating data still remain unexamined. But it comforts me to think that Kukla shares my conviction expressed in (1970, p. 103): "What should a psychological theory be? It should be analyzed—exactingly, sensitively, and exhaustively."

[5]For an insightful history of this incubus, with references to the considerable literature of efforts- all unsuccessful-to loosen its grip on psychology, see the work of Gigerenzer (most recently 1993). Also see Cohen (1995).

References

Cohen, J. (1995). The earth is round, $p < .05$. *American Psychologist*, *49*, 997–1003.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale NJ: Lawrence Erlbaum Associates.

Kim, J. (1993). The non-reductionist's troubles with mental causation. In J. Heil & A. Mele (Eds.), *Mental causation.* Oxford: Clarendon Press.

Kukla, A. (1995). Amplification and simplification as modes of theoretical analysis in psychology. *New Ideas in Psychology*, *13*, 201–217.

MacCorquodale, K., & Meehl, P. E. (1954). Modern learning theory. In W. K. Estes et al (Ed.), *E. C. Tolman.* New York: Appleton-Century.

Putnam, H. (1980). 'the nnateness hypothesis' and explanatory models in linguistics. In H. Morick (Ed.), *Challenges to empiricism.* Indianapolis: Hackett. (pp. 240–250)

Rozeboom, W. W. (1958). "What is learned?"—an empirical enigma. *Psychological Review*, *65*, 22-33.

Rozeboom, W. W. (1960). Do stimuli elicit behavior?—a study in the logical foundations of behavioristics. *Philosophy of science*, *27*, 159–170.

Rozeboom, W. W. (1961). Formal analysis and the language of behavior theory. In H. Feigl & G. Maxwell (Eds.), *Current issues in the philosophy of science.* New York: Holt, Rinehart, & Winston, Inc.

Rozeboom, W. W. (1965). The concept of memory. *Psychological Record*, *15*, 329–368.

Rozeboom, W. W. (1970). The art of metascience, or, What should a psychological theory be? In J. R. Royce (Ed.), *Toward unification in psychology.* Toronto: Toronto University Press.

Rozeboom, W. W. (1974). The learning tradition. In E. C. Carterette & M. Friedman (Eds.), *Handbook of Perception, Vol. 1.* New York: Academic Press.

Rozeboom, W. W. (1977). Metathink—a radical alternative. *Canadian Psychological Review*, *18*, 197–203.