

TO: H. Feigl
 FROM: P. E. Meehl
 DATE: December 15, 1976, (rev. 1993)
 RE: Prolegomena to post-positivist ethics¹

[This is a mostly unargued summary of several distinct but related “insights” that I have had during the last few months, precipitated partly by discussions with Dr. A. Woodard Ching (psychologist at Veterans Administration Hospital, Fort Meade [in 1976]), partly by a re-reading of Professor Donald C. Williams’ neglected 1933 “Ethics as pure postulate” (reprinted in Sellars and Hospers *Readings in Ethical Theory*), and partly by D. H. Monro’s *Empiricism and Ethics* (1967). I am also indebted to Professor Rolf Sartorius (Philosophy), Dr. Edwin L. Crosby (President, Jefferson Center, Minneapolis) and Mr. Scott Dickman (clinical psychology graduate student [in 1976]) for stimulating conversations.]

§ 1. Given the positivistic attack on the cognitive status of axiology and of ethical assertions in particular, we have fallen into the habit of assuming a meta-principle which one would not even assume in formal disciplines like logic or mathematics, and *a fortiori* not in the reconstruction of empirical science, to wit, that ‘true’ must mean the same as ‘valid.’

§ 2. Similarly, we often assume that to have truth status a statement must be decidable (Gödel sense), which is not the same as mistake (§1).

§ 3. Similarly, we often attack the cognitive meaningfulness or true-false status of an ethical statement because of the absence of an effective decision procedure, which is neither the same as mistake (§1) or mistake (§2). Remember Church’s Theorem!

§ 4. We infer from the difficulty that two sane rational persons of good will may experience in resolving an ethical dispute, and the historical fact that even within a given culture such disputes may persist indefinitely unresolved, that the sentences lack cognitive content, that they have no status as true or false, that they do not “assert

¹ This is a slightly revised version of the original text sent to Herbert Feigl in 1976 (and to a few others since), edited for grammar, slang, and updating some time-dependent references. The acronym ERDERVE—Extended Rational Discussion based on Extensive Real and Vicarious Experience—was ERDEVE in the earlier version; I subsequently added the ‘Real.’ [*Discussed in Sections 25 and 34—LJY*] Professor Feigl’s comment on this memo was, “Well, it’s a bit long for its length.” The repetitious, ruminative, hyper-qualified style stemmed from my own deep puzzlement and vacillation. I would much prefer to be an “ethical realist” or “moral objectivist,” but I find that position difficult to formulate and defend. Reading the memo 17 years later, I think it makes some valid points—especially about the need to formalize and quantify an ethical system—but today I am even less confident about the ontological status of ethical propositions than I was then. I am afraid we cannot do much better than J. L. Mackie’s *Ethics: Inventing Right and Wrong* (1977).
 —P. E. Meehl, February 1993.

anything.” This is particularly dumb of us, because the same can be said of logic and mathematics; and of course the same can be said of the reconstruction of the exact sciences, let alone the biological and social sciences. Persisting disagreement does not liquidate cognitive meaning. Remember the Goldbach Conjecture, Fermat’s Last Theorem, not to mention disputes over Bayesian inference, justificationism, the Neyman-Pearson Method, induction, minimax decision policy, quantum mechanics, relativity theory, confirmation paradoxes, Brouwer’s intuitionism rejecting non-constructive proofs in transfinite arithmetic, axiom of choice, ramified theory of types, axiom of infinity, Cantor’s Conjecture, the Loewenheim-Skolem Theorem, etc. Freud published *The Interpretation of Dreams* in 1900, yet today, three-quarters of a century later, one can easily find be-doctored practitioners and researchers who do not even agree that dreams “have a (motivational) meaning,” let alone the rest of Freud’s causal network or the specific standard symbols—e.g., snake, landscape, empress—which I as a psychotherapist find so convincing.

§ 5. We also tend to assume, in line with the Vienna positivism (and even your refurbished “logical empiricism” and “empirical realism” in which I was raised), that it is always possible, in principle, to disentangle the cognitive from the axiological or evaluative components of human experience, thought, and rational discourse. I mean “disentangle” *analytically*, of course; no one ever claimed they could be disconnected psychologically, and we logical positivists said that didn’t matter anyhow. I still wish that were so, as I know you do. But it seems to be getting increasingly difficult to reconstruct scientific knowledge, let alone other kinds of knowledge, on that “cognitively pure” basis, wouldn’t you agree? (I do not refer to Karl Mannheim’s book here, although Kuhn might think it relevant.) Even in Reichenbach’s 1938 *Experience and Prediction*, he classifies what he calls generically “decisions” into two different kinds of decisions. The first kinds, sort of cognitively harmless, are the ones Reichenbach designates as mere *conventions*. These are things like deciding how long a meter is, for instance. They make no substantial difference as to the content of our knowledge. But then, even back in those good old positivistic days, Herr Professor Doktor Reichenbach identifies a second and more interesting subcategory within the generic category of decisions, to wit, those he calls *volitional bifurcations*. The interesting thing about them is that, while they are still decisional on our part, in some sense stipulative—or even, if you want to stretch the word a bit, you could say “arbitrary,” in the sense that nobody could prove that we were in some basic sense wrong or mistaken in making them—when we do make one of these decisions of this second kind that are not mere conventions but volitional bifurcations, our resulting science will look different from what it would have been if we had bifurcated along the other prong of the decisional fork. While it’s true that the examples he gives are things such as scientist’s purposes in inquiry (e.g., studying science to fill his hours of leisure), he does make the abstract statement, not, alas, illustrated with an example, that a volitional bifurcation resembles a *bifurcation of ways which will never meet again*. Then he goes on to talk about “entailed decisions,” in terms of which the

advisory task of epistemology, its third task for Reichenbach, must be carried out. Note that this third task has prescriptive (“good advice”) elements, and so shares components with the reasoning that occurs in Reichenbach’s “context of justification.” That is, what a working scientist listens to, takes seriously, acts upon in the epistemologist’s advisory discourse cannot be summarily relegated to the “context of discovery,” since it is (a) *prescriptive discourse* that, if heeded, (b) *makes a difference in the content of our scientific knowledge*.

Whether Reichenbach vaguely anticipated them or not, I shall simply *list* (not expound) some of the more recent developments (which, I confess, go against the grain for me because I dislike to see the “purely cognitive” infected with axiological elements) to give you an idea of what I have in mind. Incidentally, I don’t think any of these have been cooked up by people for obscurantist purposes. I believe the record shows that each of them was, so to speak, stumbled upon and even, perhaps, reluctantly admitted by thinkers in the broadly analytic and empiricist intellectual tradition, thinkers who are not grinding some kind of Marxist or Christian or Hegelian or Fascist or other ideological axe. I have in mind such stomach aches as the following:

- a. The problem of cognitive-aim decision theory, i.e., of choosing an acceptable decision policy as regards *theoretical science*. I include here the paper by John Milnor showing that no intuitively satisfying consistent decision system has been formulated. We write down some simple, “obvious” *meta-meta-rules* for choosing a decision policy, each of which, taken singly, has great intuitive appeal to reasonable minds. Then Milnor shows that we get nine such intuitively plausible “necessary” criteria for a satisfactory decisional policy, and that none of the four currently available (expectancy, minimax, Hurwicz, least regret) satisfies them. He leaves us with a query whether it is even possible to formulate a decision rule that would satisfy all nine.
- b. The problem of assigning Bayesian priors in theory testing.
- c. The problem of arbitrariness in the choice of significance levels for determining Type I versus Type II errors in statistical inference. As we lower the significance level α (say from $p = .10$ to $.01$) hence decreasing our rate of Type I errors, we unavoidably reduce power $[1 - \beta]$ and increase the incidence of Type II errors. In quality control the balance of $\alpha : \beta$ can, in principle, be nonarbitrarily chosen by reference to costs (e.g., angry customers, dead patients). In theoretical science no one has shown how to do such a thing, and many believe that it could never be done.
- d. The still unresolved problem of why a maximum likelihood solution should be chosen. R. A. Fisher never proved, and I understand from my statistician friends that nobody could possibly prove (without illicit Bayesian assumptions that get us into the traditional difficulties when Bayes’ Theorem is misused), that choosing

the parameter value which makes the observed point in sample space more probable than any other parameter value would make it, is the “best choice.”

- e. Problem in (d) of even saying clearly what we *mean* by “the best choice.”
- f. Problem of Popperian risk-taking in one’s theorizing as contrasted with Skinner’s inductivism (I would have in mind here the difference between people who like interesting strong theories vs. people who prefer less interesting “safe” ones that are closer to the facts).
- g. The problem cluster of Lakatosian problem-shifts, degenerating research programs, theoretical tenacity, Feyerabendian advocacy of “continuous maximum proliferation” of theories—that whole can of worms that’s involved, *even within a purely theoretical ballpark* in which the only stakes are truth and error, in what is a rational way to go about doing empirical science.
- h. The problem of Kuhnian revolutions, normal science vs. challenging the paradigm, and the like.

Most of this stuff one finds in such books as Lakatos and Musgrave *Criticism and the Growth of Knowledge*, Levi’s *Gambling with Truth*, or the new one (fascinating!) by Nicholas Rescher on *The Primacy of Practice*.

§ 6. We seem to assume that a crude rule of thumb way of classifying complex circumstances must permit resolution of all ethical issues, when we have not yet finished—in fact, have barely begun—building the formalism of such disciplines as decision analysis, utility theory, deontic logic, and the like.

- a. This is a very important point—actually the main point motivating this memo—because it is possible that ethical reasoning about complex cases, as carried out even by sophisticated persons at the present time, is in an extremely primitive state. “Ethics 1975” is comparable to medieval physics, or Egyptian (as contrasted with Greek) geometry, or medicine at the time of Paracelsus. I see here an important analogy to Hume’s treatment of natural science (which he apparently didn’t know much about, being a historian by trade). Hume’s empirical examples of alleged natural laws, natural necessities, “causes” and “effects,” are usually quite simple-minded prescientific examples that most of us wouldn’t today consider laws of nature at all. He considers igniting fires with matches, thunder following lightening, dropping stones on glass, crows being black, and the like. Nobody (except a few humanistically trained and scientifically ignorant Oxbridgeans) believes that such examples, or the explanation-schemata fitting them, can be useful for a philosophical understanding of scientific causal laws or for giving the scientist helpful methodological advice.
- b. Suppose it is true, as I suspect, that something as complex as Robert Nozick’s brilliant paper on “Moral complications and moral structures” (*Natural Law Forum*, 1968, 13, 1–50) is merely the beginning. (Or see, for another example,

Carlos Alchourrón and Eugenio Bulygin, *Normative Systems*. New York-Wein: Springer-Verlag, 1971.) Then the fact that two informed, rational, and humane persons can sometimes have different ethical intuitions about a complex case (such as court mandated school busing, selective black admissions, the Reserve Mining pollution case, shock therapy in manic-depression, capital punishment for kidnappers, or the graduated income tax) does not show, or strongly tend to show, what we positivists usually have taken it to show, namely, that the underlying *reason* for the unresolvability is simply that ethical sentences have no cognitive content anyway, that they are neither true or false in any literal ontological sense, that there is no extra-linguistic domain of entities to which ethical terms refer and to which they can possibly “correspond” or not (i.e., be either true or false.)

§ 7. A (I am tempted to say ‘*The*’) short answer to the hard question, “What is the *realm of being* [I’ve been reading Santayana lately] in which ethical predicates obtain?” is this: “They belong to the empirical order of human beings and infra-human animals, angels (if there are any), and disembodied spirits (if there are any).” Only a moral agent can be a value of the first variable appearing in an ethical relation. Only another moral agent, or another sentient being, can occur as the second term in such a relation. Entities that are not moral agents and not sentient beings (such as stones) can only enter such discourse via relations alleged to exist between the others, as we know in jurisprudence. Thus, for example, ownership of property is not literally a relation between me and the boulder on the farm that I own, but is a relation between me and other people as regards what we can and cannot do with or to or on the boulder. Successful outcome of a quiet title action is a declaration by the court of my rights “against the whole world,” as the lawyers put it—not my “rights *vis-à-vis* Blackacre,” the latter not being a legal person or a moral agent. One has to include both moral agents and sentient beings, because we don’t want to exclude the possibility that a person might have obligations to somebody solely because he made a promise to another moral agent, quite apart from whether the latter was capable of suffering pain or pleasure or whatever. I doubt that, but I want to leave it open. I certainly want to say that we have obligations to sentient beings that are not capable of being moral agents. Thus, for instance, one of my obligations is to avoid cruelty to an infrahuman animal. I believe that we have obligations to such animals, but that they are not capable of having obligations toward us. That is also known to be an arguable issue, but again I don’t want to foreclose it.

§ 8. It sounds pretty spooky, and I am not at all confident that I can defend it, but I want to argue that whenever certain “descriptive” relations obtain between a moral agent and a sentient being, then *ipso facto* such-and-such “ethical” relations also obtain between them. The ethical relations do not *consist of* the descriptive relations (that is, we avoid Moore’s “naturalistic fallacy”) but the ethical relations follow directly, via our ethical “pure postulates,” from the descriptive relations. The ethical predicates are “supervenient,” I believe Moore calls them. And the ethical relations are *objectively there*, whenever the requisite descriptive relations are there. Suppose somebody persists

in saying that he doesn't understand the ontological *locus* of ethical relations, despite my saying loud and clear that the entities which they are "about" are the empirically familiar entities: moral agents (domain) and sentient beings (co-domain). (READ: If you don't believe in angels or devils or disembodied spooks, the class consists of people and infrahuman animals.) I would reply that we have to deal with a distinctive, strange, unique kind of relation which we call 'ethical.' I would have to argue that such relations between biological organisms (such as two persons on a raft, or in a business partnership, or in a marriage, or playing chess with each other) *come into axiological being* when these two persons are in certain kinds of actual or possible causal connections with one another, however remote. It is a mistake to look for a natural property or relation like weight or color, or even IQ or social dominance, as being the ethical relation between them; and in this much I agree with G. E. Moore. Nevertheless, the ethical relations *depend upon* the existence of certain natural properties and natural relations, such as, that one person is the biological parent of another one and hence has an obligation to nurture the child; or that one person made a promise to another one and, hence, had an obligation to perform. But while ethical relations *come into being* by virtue of the existence of natural property type relations, even including such simple ones as physical position (I should not step off one end of a board if that will causally bring it to pass that you, being at the other end, are precipitated into an abyss!), they are nevertheless not *conceptually reducible* to nonethical predicates and relations. That is why Williams' notion of "ethics as pure postulate" comes into the picture. There is a distinctive ethical primitive relation or predicate, and it occurs in primitive propositions that relate the ethical predicate to naturalistic nonethical predicates and relations. So, in other words, you have to look at the physical, biological, psychological, and even economic and political "natural predicates and relations" that obtain between moral agent *X* and sentient being *Y* in order to ascertain what the distinctly ethical relationships are that obtain between them. But having done that, you realize that the latter kind of relation is not reducible to the former. Rather it *depends upon* it, via the distinctively ethical postulate, whatever that may be (e.g., "it is obligatory to avoid cruelty").

I could make an analogy (like all analogies it is imperfect but nevertheless, I think, somewhat illuminating) between ethical relations and other nonsimple but nonethical relations. Suppose we have these two fellows on a raft following a shipwreck, and they are both of the same age, unmarried, in excellent health, neither of them was injured in the shipwreck, and they have a loaf of bread and jug of water on the raft with them. They have reason to think that the SS Carpathia is on the way, it having responded to the SOS before their own ship sank; but they don't know for sure, and they don't know how long it's going to take. Query whether Smith should drink all the water and eat all the bread and then, when Jones becomes weak from thirst and hunger, push Jones off into the water? Now I maintain that any sane rational person would agree that they should share it, so if one of them eats and drinks all of it and the other one perishes, that is an unjust arrangement. This is a distinctively axiological judgment. Nevertheless the ethical

diagnosis clearly hinges upon the “naturalistic” (descriptive, non-axiological) properties of the situation. Further, I would hold (with Skinner, Ralph Barton Perry, and the “legal positivists”) that we can *define* such value-loaded terms as ‘justice’ or ‘equal treatment’ in non-axiological psychologese, provided that the latter’s vocabulary is sufficiently rich (i.e., psychophysiology + phenomenology + behavior dispositions). Nevertheless I hold that there are irreducible ethical postulates, because the naturalistic definability of the *term* ‘justice’ does not lead to a naturalistic derivability of the *proposition* “One ought to do justice.” How many ethical postulates we need I leave aside.

Suppose somebody still wants to know what is the ontological status of all this? We tell him that the *relata* entering into this ethical relationship are these two human beings. We say that they have a position in space-time, and possess certain naturalistic properties and relations, including the property of being able to experience suffering and the biological ability to die and the like. But then suppose somebody insists that he still doesn’t understand “where” the ethical relationship *exists*. We might reply that “where?” is an inappropriate question, at least if simplistically interpreted. The question implies (taken crudely) that we should be able to *point to* some kind of physical object located midway between them on the raft. But when we say (in terms of the earth’s geometry) that Jones is directly northwest of Smith on the raft, would we look for some kind of a material object, like a “northwest pointing bar,” connecting their two bodies? Or, to make it even more complicated (but still, I suppose, basically naturalistic—this is not clear since I don’t know how much of a naturalistic analysis is possible in jurisprudence), we might ask about the two men on the raft which of them is wealthier. That issue involves reference to certain documents that can be found in the First National Bank of Minneapolis and the Third State Bank of West Overshoe, together with some very complicated legal and economic statements about the value of a such and such stock in International Silklined Casket Corporation at the close of today’s market. I don’t know whether one can give a thoroughly “naturalistic” analysis of the concept of *ownership*; therefore I don’t know whether I can do it for the concept of *wealth*. I assume one could do it, but some think not. But nobody is so naive as to think that we can answer the question about wealth by looking for a “simple” physical property or, unlike the northwest proposition, even a simple “physical relation” between the two. (If, however, someone urges that the “location” of an ethical relation between two persons can at least be said to be “where they are,” I would make no objection to a suitably framed convention of that kind. *Cf.* the conventions about the vector representing gyroscopic force, or Aquinas’ “an angel is where it acts.”)

My point in all this is that *relations between persons* do not have to be relations between two simple properties, such as “Jones weighs more than Smith does by 37 pounds.” There are *nonethical* relations which are complicated to unpack, such as “being northwest of somebody,” or “having more money in the stock market than somebody else.” I am not saying ethics is no different from these. On the contrary, my view is that at least one ethical predicate is unique and not reducible to economic or social or

psychological ones. I am merely stressing that there are complicated relations that can exist between two people on a raft that do not even tempt us to commit Moore's "naturalistic fallacy" by looking for some simple predicate like weight or size or color or even physical position on the raft or how hungry Jones is in comparison with how thirsty Smith is. And my aim in stressing that is merely to "soften you up" to considering seriously the notion that *distinctively ethical relations* may be relations between biological organisms and may *depend upon* certain naturalistic predicates and relations, without being completely *reducible* to them.

§ 9. The sense in which ethics is pure postulate is the sense in which at least one primitive predicate, namely *obligation*, is introduced and is not reducible without loss of meaning to so called "descriptive" predicates. That is, we avoid any possibility of committing the naturalistic fallacy of G. E. Moore (which he shouldn't have called that). We avoid this error by stating one or more primitive propositions that involve at least this one primitive axiological predicate, combined in these primitive propositions with non-axiological predicates.

§ 10. I think of *obligation vectors* as coming into being under stated conditions as functions of the values of descriptive functors and predicates involving the relation of moral agents to sentient beings; and I think of a formalism (such as the decision theorists and econometricians and others will continue working out in ensuing years) for vector combinations of obligation vectors; and finally I think of a rule of action that says something about what one is obligated to do given a certain *net obligation vector*. Thus, for instance, it may be that, whatever the pattern of the component strengths, if the length of an obligation vector is smaller than a certain amount one might have a decision rule that says that to act this way is licit but not obligatory [= "optional"]. There are many such quasi-neutral situations in ordinary life. The formalism of such a system would contain rules that correspond to cruder results of the arguments among the seventeenth century Roman Catholic moral theologians (rigorists, probabilists, and probabiliorists.)

I suspect that an axiomatized and numerified, enriched, multiple-postulate, *and econometricized* axiology would be able to dispense with what have been called "preference rules," used (in metalanguage?—unclear to me) for decision-making when two first-level obligations conflict. That is, I conjecture that the main (only?) reason for this is that the *prima facie* obligations are customarily stated qualitatively, rather than in a quantitative ("ethical vector") formalism. Even lexical orderings à la Rawls can be rendered econometrically, as I understand it.

§ 11. In the light of these considerations it should not surprise anybody that the ethical intuitions of informed persons tend to converge for simple cases like the two men on a raft but diverge increasingly with apparently irresolvable disagreement as we move to more and more complex cases. I urge that no meta-inferences as to the meaningfulness or truth value of ethical statements be drawn from this, contrary to what the positivists (myself included) have been very prone to do.

§ 12. When I say, “Hitler ought not to have murdered six million Jews” and allege that this follows from some system of empirical statements together with one or more primary ethical postulates, and then I tell you that I look upon my primary axiological commitments underlying this moral judgment as “pure postulates,” I am *not* thereby admitting that they have no validity or truth status. It is misleading to analogize this kind of moral example to “I prefer vanilla to chocolate malts,” via a meta-ethical statement that says “Primary axiological statements are merely expressions of personal taste.” All four of these words—‘merely,’ ‘personal,’ ‘expression,’ ‘taste’—heavily load the dice against the ontological status of ethical predictions. Of course that I prefer vanilla to chocolate malts can be false because it can fail to accord with my introspections if for some reason I am careless or desirous of deceiving you, as for tactful reasons if I get stuck with the vanilla one and I know you strongly prefer chocolate. It may be possible, if there is an incoherency between my best available introspected preferences and my choice behavior, that somebody could show me that I tend to choose chocolate more often than vanilla; and that finding would then present a problem of explanation. Of course it doesn’t immediately refute the former, because I might be a masochist, or I might be an anal character who deliberately denies myself pleasures so that in some sense I will deserve them more when I do get them; or maybe I believe that the pleasures feel better when I get them more rarely, as some anal characters are even able to report consciously when asked in psychotherapy.

§ 13. The interaction between concrete intuitions and generalized axiological systems is analogous to that between single observations (which we think of as somehow legitimating a language entry move, that is, the tokening of a protocol statement) and the structure of causal theory. It is misleading to say that the facts do not have anything to do with the theory or do not in some sense ultimately and collectively control the theory; but it is equally misleading to deny that theories ever control (individual alleged) facts. It is perfectly obvious, both in the history of science and in the practice of the scientific laboratory (and in common life), that it works both ways. Some complicated reconstruction must be given to do justice to the works, and nobody has done that reconstruction yet.

§ 14. My unpublished paper “Are we all crypto-Prichardians?” needs to be re-examined in the light of these considerations.

§ 15. I leave open how many primitive predicates we require, not explicitly reducible to “naturalistic-descriptive [= non-axiological] predicates.” My hunch is that we need only one, namely, ‘(...obligated to...)’. But possibly we need two, the other being Moore’s ‘(...good)’. Others, such as ‘cruel,’ ‘fair play,’ ‘dignity,’ ‘restitution,’ ‘promise-keeping,’ ‘merit,’ ‘share,’ are, I think, all reducible in principle to psychologese and economese. However, even if there’s only a single ethical primitive *concept*, there may be several primitive (“pure postulate”) ethical *statements*. That is, we may have several ethical predicates that are reducible to psychologese (such as ‘promise’ or ‘cruel’); then we have one ethical primitive, such as ‘obliged to’; but this setup gives several primitive

ethical postulates, such as ‘I am obliged to keep promises,’ ‘I am obliged to avoid cruelty,’ and so on. (I put the moral primitives in conventional qualitative form, since I do not know just how to write them econometrically.)

§ 16. It seems plausible to conjecture that when an adequate complex axiological econometric formalism is developed (I shall hereafter speak of *axiometrics* = axiological econometrics), what appear currently to be ethical questions will sometimes turn out to be computational questions and frequently will turn out to be empirical questions in the ordinary descriptive sense of that word. This much I give the Skinnerians gladly. Example: The ethical dilemma ‘‘Should I beat a child to help him learn his catechism?’’ as was standard procedure in the days of Luther, would have required an ethical sophistication which could only have sprung (except in an unusually saintly person) from a psychological developmental sophistication about the effects of certain kinds of handling upon the child mind and his subsequent adult problems. We cannot infer that Luther’s schoolmasters were all cruel and ethically dull persons because the rod was used as part of the shaping up of the child’s behavior in Latin or arithmetic. They literally didn’t *know* enough about psychology and child development and so on for this ethical problem to have presented itself in the way it does to an informed contemporary mind. Similarly, Kant’s views about universalizability must be examined today using knowledge of the effects of crf versus VR reinforcement schedules.²

§ 17. Even if the general line of reasoning were correct, one still has the uncomfortable feeling (attributable to what I now view as my positivist prejudices, but maybe there are really good reasons) that there is an ‘‘arbitrariness’’ involved here that is not involved in choosing the postulates of logic, nor in selecting competing empirical theories in the factual sciences. Several related responses occur to me in response to this complaint, and taken together they may provide at least a feeble reply. At least they suggest a direction for our thought to move in.

The first response is to reiterate that in the empirical sciences an element of choice or decision exists in most sciences at most stages of theoretical controversy, and that truly decisive hammer-blow experiments are rather hard to come by. But suppose somebody says that, nevertheless, a point is finally reached in a clearly degenerating program (*pace* Feyerabend!) where only some kind of a nut would stick with a thoroughly refuted empirical theory. To this I reply by asking whether we now know that *a sufficiently rich, formalized, interknit axiometrics would be so much worse off in that respect*, assuming we ever come to have such? Imagine a sufficiently enriched, formalized, axiomatized and quantified axiological system that would handle such diverse cases as the two men on the raft, chattel slavery, and deciding who should suffer the loss in a bad car accident (in

² [Examples of schedules of reinforcement that produce different behavioral outcomes; a crf (abbreviation for continuous reinforcement) schedule reinforces every target response, whereas a VR (variable ratio) schedule reinforces the target response some average percent of the time. See C. B. Ferster & B. F. Skinner, *Schedules of Reinforcement*, (1957).—LJY]

terms of whose “fault” it was, or whether we want to have no-fault insurance for other reasons than to do justice), and so forth. Then it might turn out you would have to be almost as much of a moral nut to choose Hitler’s (rather than Buddha’s or Christ’s or John Dewey’s or Prichard’s or Kant’s) ethic as you would to insist on the flat earth theory in the present state of the evidence. I remind you that there are still people who belong to an outfit called The Flat Earth Society and who claim to explain away even the visible roundness of the earth as seen by the astronauts! Or, to be less crazy, we know that some competent astrophysicists believe that Einstein’s theory is false, despite all its evidence, that it has high verisimilitude, but still is radically false, e.g., that c is the top (“physically possible”) velocity. In a way what I’m saying here is rather trivial. It amounts to the familiar point that if somebody is screwy enough he can stick to almost any theory despite conceptual and factual difficulties, and that this human possibility is not a hammer-blow objection to the ontological status or to the intrinsic “objectivity” of the properties and relations that a certain domain of knowledge deals with. It *looks* as if it is a lot more arbitrary in ethics than it is in physics. But if I am right in thinking that we try today to deal with ethics as Hume tried to deal with natural science and if I am right in supposing that a thoroughly rigorous formalization, comparable to that of the top level econometrics and decision theory stuff already being done, will some day be done for ethics, then the arbitrariness might be very greatly decreased. It will still be true that somebody can ask whether we are requiring the system to fit our intuitions about the concrete case, or the other way around. I have tried to deal with that in this memo, although briefly and dogmatically, by pointing out that *it works both ways*, just as is the case in physical science, or even in the formal disciplines. An axiological system that did gross violence to *all* of our concrete ethical intuitions, even in very clear-cut instances such as the two fellows on the raft, would not be tenable, and we would try to give some kind of (doubtless complicated) meta-account of this untenability. I suggest that one reason why we cannot give such a meta-account at present is that we don’t have the elaborated object language axiology that I envision. I repeat: To some degree our concrete intuitions for clear cases impose constraints upon what is an acceptable general axiology; and yet also, and not contradicting this remark, we have the other derivational direction going, where some of our more doubtful intuitions about the concrete are influenceable by the application of an otherwise well supported general axiology. So I urge you to consider that the relationship between facts and theories, or between the desired theorems in a formal system and the intendedly complete axioms that will generate them, is, while different, not utterly unlike that which would exist in a well developed ethical system.

However, one might say that even given constraints that we can’t presently visualize until somebody builds such a system for us and we can try it out in a variety of situations, wouldn’t there be a residual ambiguity of a quantitative nature? Somebody might postulate fundamental axiologicals generating obligation vectors based on mixed considerations of justice and of utility, such that sometimes there would be quantitative com-

promises between total human happiness achieved over all persons and the insistence upon avoiding injustice to individuals, even though such injustice might benefit so many more other people that the net human happiness would be increased (e.g., Aristotle's views about slavery). That's a very bad one and I do not have what I consider at all a satisfactory answer to it. But I give you this one thought, which I think deserves more detailed attention: Suppose we said that, just as there is a certain intrinsic "coarseness" in the physical order (at least since the development of quantum theory and probably before then), and there is an allowed crudeness or looseness in concepts used in certain biological fields such as clinical medicine, might it not be the case that an adequate meta-theory concerning axiological systems would allow for such a coarseness? We would say that, just as some net obligation vectors have such a short length that whether one performs an act in the direction indicated by the vector or fails to perform it is to be considered ethically neutral (that is, it is licit to perform it, but it is not required), so we might say that it is an intrinsic feature of value systems that they have a rather large "quantum uncertainty" built into them, so to speak. Perhaps it is not reasonable to require of an axiological network that it should always decide clearly between the different intuitions of Jones and Smith for all possible gray region concrete cases. Putting it crudely, one might say, "There are moral differences that are so slight that they don't make a difference and only a moral pedant or perfectionist or somebody running for chief saint would worry about them." I am suggesting that there is a "fineness of grain" in applying an idealized axiology beyond which we would not pursue questions of validity or preference, just as in working out the mathematics of tax economics we might use continuous functions even though we know that the smallest amount that any taxpayer ever pays is whatever is the smallest unit, such as a cent or a mill. I don't know whether this will do it, and I suspect it won't quite take care of an element of "primary commitment" that is hard to understand in terms of objective truth value. But I think there's something here worth pursuing.

§ 18. I want to avoid any appearance (as I am pretty sure I have avoided the substance) of a *tu quoque* fallacy. Ever since I took beginning logic I have been a bit hazy about when "Your nose is snotty too, Socrates" is a fallacy and when it isn't. Whereas this memo adumbrates a constructive, positive job that I hope to do one of these days, it does not do more than adumbrate it, and its main point is to attack and destroy a batch of positivistic and related anti-ethical meta-criticisms. Ethics as "pure postulate," along with the detailed quantitative decision-theoretical mathematics that one would need in such an ethics to do the "cut-and-fit" job between the theory and the very strong concrete ethical intuitions (upon which someone like Prichard or Ross relies) is music of the future. I do not claim here to have done more than the feeblest initial sketching of a positive meta-theoretical account of the ontological status of ethical relations. Notice that I say 'relations,' not 'entities,' because the *entities* are human beings and other sentient creatures to whom they have obligations. But I do think to have shown, pending rebuttal of my arguments, that those of us in the broadly positivist tradition of thought have

permitted ourselves anti-ethical arguments, arguments purporting to show that distinctively ethical expressions *could not have* any objective ontological status, *could not have* even any cognitive meaning, which meta-arguments we would *not* apply across the board. And that kind of a “Your nose is snotty too, Socrates” position seems to me perfectly respectable philosophical reasoning. A positivist says that ethical consequences flow from primitive propositions that are not themselves proved but are “pure postulates” (in some sense); that it is unfortunately true that intelligent and humane and reasonable men can have persisting ethical disagreements about complex cases; that we do not have a perfectly precise and clear and indisputable metatheory about the correspondence theory of truth for ethical statements; and that some ethical judgments fall in a gray region which even an axiomatized and numerified quantitative theory of obligation vectors might leave us in doubt about. To all of these complaints, we reply that yes, they are perfectly true about ethical reasoning. Are we then entitled to assume with our positivist critic that any beliefs or judgments or statements which suffer from any of those defects I have just mentioned are automatically thrown into the ash can as noncognitive, asserting nothing, having no reality status, no truth or falsity? I say not. Because, we go on to argue, *if* he is committed to the view that if a sentence has a truth value it must have a finite decision procedure available, or that a sentence cannot have truth value if it occurs as one of the postulates of a system rather than as a theorem that is proved therein, or that a sentence cannot have any truth-value or cognitive content if intelligent and humane and informed and reasonable persons seem to persist in disagreeing about it even after lengthy discussion—then of course, if my reasoning is right, he will have thrown out physics and chemistry and obviously all of the biological and social sciences, and he will probably have to throw out most of mathematics. For instance, he will have to throw out set theory, which is considered the most fundamental and general kind of mathematics, inasmuch as Paul Cohen has shown (am I mixing him up?) that the Cantor Continuum Hypothesis is neither a theorem flowing from the von Neumann axiomatization of set theory, nor does the addition of the Cantor Continuum Hypothesis generate contradictions. Therefore, the Cantor Continuum Hypothesis cannot be decided on the basis of the axioms of set theory, from which some meta-mathematicians have inferred that it is neither true nor false. Whereas others have said that this is merely another instance, comparable to what some people have said about the Loewenheim-Skolem Theorem, that a theory of meaning for even abstract entities that is based *entirely* upon the notion of “implicit definition” just won’t wash. That is, some people are saying, if I understand the controversy aright, that the Cantor Continuum Hypothesis is either true or false, that there either is or isn’t a transfinite cardinal lying between the cardinal number of the integers and the cardinal number of the points on a line, and that we are not entitled, absent a proof, to identify \aleph_1 with C . Now I am aware that people have used all sorts of problems in physics and mathematics for metaphysical and theological obscurantist purposes, but I don’t believe I am doing that. What I am saying is that anybody who throws out the cognitive status of distinctively ethical sentences or systems on certain

kinds of metatheoretical grounds ought to be able to give an account of why he believes in mathematics or logic or physics or behavior genetics or Freudian psychology or Keynesian economics, despite the fact that the meta-grounds he is using for liquidating ethics would also liquidate these other cognitive enterprises which he believes have cognitive meaning and, in some sense, “objective” truth or falsity.

§ 19. There remains in my mind, as I am sure in yours, a deep, pervasive, and perhaps intractable problem about the ontology of axiological statements, given that you and I both subscribe to some funny mixture of the correspondence and coherence theories of truth. I am very much afraid (I say “afraid” because, like you, I strongly disapprove of Hitler and Stalin, but I don’t disapprove of somebody’s preferring Schoenberg to Telemann [although I wouldn’t pay a nickel to hear most of Schoenberg—on the contrary, I would gladly shell out ten dollars to be free of having to listen to him], and therefore I find the idea morally “fearful”) that even when (as I am confident will happen) the next generation of young economists and philosophers succeeds in developing a rich, subtle and rigorous formalism for utility theory and deontic logic; and when (as I am equally confident will occur) we have developed a powerful science of behavior and phenomenology free of some of the current “school” prejudices, these developments will make possible, I hold against writers like Friedman and Hayek and Mises, meaningful interpersonal utility comparisons so we can dispense with Pareto optimality—one of the most overblown and useless concepts ever seized upon by social science. Yet, given all this, I very much fear that there will be persisting disagreements among intelligent, informed, rational, humane, ethically perceptive persons which will involve not tiny but sometimes moderate to large differences in the net obligation vectors, derivable from differences in the quantitative parameters of their primary ethical postulates. I may be wrong about this, and I hope that I am. I don’t see how anyone can be dogmatic about what an axiomatized and numerified, highly formalized mathematicized ethics, when conjoined with an advanced behavioral and phenomenological empirical science, would look like. We cannot say which among the conflicts that currently appear irresolvable in the guise of rock bottom axiological collisions would, under such advanced knowledge circumstances, turn out to be somehow resolved. There have been too many examples in the past of people philosophizing with great assurance about matters whose resolution nobody could foresee *because nobody realized what empirical science and mathematics and logic might come to do*. Take one of your favorite examples, the positivist Comte asserting confidently that one thing mankind couldn’t ever know, in his sense of “positive science,” was the chemical constitution of the stars, because they were not available for test tube inspection. There was no way for Comte to foresee the crazy possibility that one could ascertain the chemical constitution of a distant fiery globe of gas by analysis of the distribution of light waves transmitted from it to the earth, such that today we know the percentage chemical composition of the sun more accurately than that of the earth itself!

Nevertheless, I don’t have much faith in my own hope for such results in the ethical domain. I am afraid that a combination of a sufficiently rich and rigorous formalism of

ethical “pure postulates” on the one hand, with a sufficiently well corroborated and detailed psychology of behavior and experience on the other (I include here the ethically relevant features of econometrics, of course), would find persons ultimately disagreeing in important amounts in the light of all the evidence. Consider, for instance, the threefold motto of the French Revolution. It seems pretty obvious that humane, intelligent, rational minds do differ quantitatively in how strongly they value liberty vs. equality vs. fraternity, even though they may (qualitatively) value all three. It is a sad fact of society that sometimes in order to get more of one of these three we must put up with a decrement in one of the others. Thus, for instance, I don’t have much doubt that however society comes to move in the future (even, as I would favor, to some sort of nonstate cooperative socialism) it will be impossible to maximize or optimize or even, to use Herbert Simon’s word, “satisfice” some extremely passionate equalitarians without dissatisficing the basic value commitments of extreme libertarians. As you know, I myself find these three French Revolution values strongly ordered in my own mind. I am a fanatic on liberty, I give a second place (and except for income in the ideal society, a rather poor second place) to equality, and I couldn’t care less about fraternity. In fact I am actively opposed to any efforts by Big Brother to impose fraternity—racial, class, religious, student, or otherwise—upon me or upon anybody else. The Minnesota Civil Liberties Union, to which I belonged and of which I was a fervent supporter until rather recently (when I found that I had “flunked” their questionnaire, not because *I* had changed but because *they* had changed), was active in pushing through legislation requiring the State Liquor Commissioner to refuse to issue a liquor license to a private club, such as the Elk’s Club, if it has either a rule or an informal policy of restricting membership to Caucasians. Now, I would never join a social club that had any sort of racial bar, either by its official rules or by a tacit understanding. If I found that I had inadvertently done so, I would try first to reform it from within by persuading my fellow members of the irrationality and inhumanity of such a policy; and then, if that didn’t work, I would resign my membership in a highly visible way (with carbon copies of the letter of resignation and the reasons sent to suitable public media, etc.); and next I would proceed to try to dissuade others who contemplated joining the group from doing so until such time as it ceased to practice racism. I think my own record on this matter is fairly clear, granted that we are a “racist society” and that I am a product of it. Nevertheless, I am strongly opposed to this legislation, which the Civil Liberties Union was proud of having helped get passed, because I consider that when it comes to a private social club, if Archie Bunker doesn’t want to drink beer or play pool with blacks or, as in his case, Italians or Catholics or “Polacks” or, if I understand Archie’s attitudes, practically anybody but North Europeans (how about the Finns?), that is Archie Bunker’s right. I think Archie Bunker is a bigot and a fool and an indigenous fascist. But so long as he is not dean of a medical school discriminating in admissions, or engaged in a public business such as that of inn-keeper or a bus line or a public utility, or the director of a city school system, I maintain that the liberty of a person to be a racist bigot, the fundamental right of

choosing with whom one wishes to associate socially and sexually in a private context—and that includes the right to pick them on whatever crazy and ill informed basis you please, such as only liking members of the Royal Order of Hibernians who tell their beads daily or preferring Norwegian Lutherans who frown on beer and cigarettes or people who practice Zen Buddhism or hard-core atheist communists—all of these personal preferences are in my opinion silly, but they are still part of Archie Bunker's (or anybody else's) human rights. In other words, I do not accept the concept that the "right to associate" is unilateral. The statute requiring the Liquor Commissioner to refuse a liquor license to the Elk's Club because it is racially discriminatory is in my opinion an abuse of the state's police and health power. I am aware that appellate courts in such circumstances are usually liberal with the legislative intent, holding that if the legislature could "reasonably have concluded" that such and such a constitutionally permissible social aim was significantly helped by the statutory provisions, that's enough. But I am not talking about constitutional law here. I am talking about the ethico-political pressure of the Civil Liberties Union in getting the legislation passed. Some appellate courts might say that this is an improper use of the police and health power, because it would be pretty hard to argue that the social purposes of a liquor license, which are mainly to keep people from drinking bootleg whiskey or methyl alcohol that will make them go blind or to prevent somebody from catching gonorrhea because there's a whorehouse upstairs and the like, are significantly improved by requiring the Elk's Club to admit blacks or Chicanos or Amerindians or whatever! But, I repeat, I'm not mainly concerned with the strictly legal side of it. I am concerned with the fact that nobody with the slightest familiarity with the legislative history or the motivations of the MCLU leadership has any doubt what is actually taking place, to wit, the police and health power of the state is being used as leverage to coerce the Archie Bunkers in the Elk's Club to socialize with people that they prefer not to.

My hunch is that even a very fairly elaborated ethical formalism, combined with a very thoroughly elaborated science of behavior and experience, would leave someone like me, who is personally opposed to racism and would never join the Elk's Club because of that fact, in a rock bottom collision with somebody like Lynn Castner or Matthew Stark, because it is fairly clear that our basic ethical postulates are different in a significant quantitative amount. That is, they believe that the concept of liberty (or is it equality?) includes the right to associate against somebody else's will in a private social club. They also believe that the state should get into the act in enforcing fraternity, even if that means, as it clearly does in this instance, a major impairment of Archie Bunker's individual liberty of association. And my hunch is that neither the developed formalism, nor the expanded empirical knowledge of mind and society which I envisage, would resolve that kind of quantitative difference. I think that somebody who is a fanatic on autonomy (as you know, the best way to make me angry is to start coercing me to do something that I don't find it interesting to do and that I don't believe I have a moral or legal obligation to do), and who gives equality (except for monetary equality, which I

favor for Abba Lerner's reasons) a moderate to poor second weight, and who gives fraternity an extremely low weight—so low that I consider it essentially none of the state's business to coerce me or Archie Bunker or anybody else into it—would have persisting disagreement.

I have said earlier in the memo that I don't view persisting disagreements as fatal to claiming some sort of objective ontological status for ethical relations, since we have persisting disagreement in philosophy of science, not to mention within the sciences themselves, including the so-called exact sciences on some issues. However, it seems to me that there is some sort of important difference between a persisting disagreement which one has some notion of how he *might* imagine himself resolving, and one where it seems hard to even conceive of any sort of empirical or formal steps that would be able to resolve it. If I am a liberty nut and you are an equality or fraternity nut in basic passions and commitments, are there any *facts* that could help us resolve a strong quantitative difference here? I doubt it very much. But still—might not a rule-utilitarian ethical calculus, combined with a much more detailed and precise psychology than we possess today, lead me to revise my libertarian priority? How could I dogmatically say to such a question, "No, it could not possibly, ever"?

§ 20. In the last paragraph of Section 17 I mentioned a certain "coarseness" that might be found ineradicable in even an advanced (formalized, axiomatized, reticulated, econometricized) utopian axiology. A related point is suggested by the way in which a good Skinnerian like MacCorquodale deals with hypothetical questions that students put to him about "facts" that would seem to contradict the system. A student will say to him, "Well, how would Skinner explain what happens if..." and then spin out some sort of conjectured business about behavior which he imagines *could* or *might* take place under such and such imagined circumstances, but which is not an account of anything that he has actually observed under adequately specified empirical conditions. This is a very tempting sort of problem for most psychologists, and one thing that the Skinner position does for those who use it consistently and with conviction, like MacCorquodale, is that they don't have to buy this type of challenge. MacCorquodale says to such students, "Why should I try to explain in Skinnerian terms how something *could* happen, when I have no good reason to think such things *do* happen? I am not in the business of explaining hypothetical occurrences which I doubt take place; and why should you want to do this, or expect that Skinner ought to try to do it?" I think that with regard to the "horrible hypo" kind of refutation so much relied upon in ethical argumentation, sometimes the axiologist should take a similar hard line. Just as it is the purpose of causal nomologicals to *subsume observational facts* (therefore we criticize a nomological network because it fails to include a well corroborated observational result), so the business of axiological networks is to *prescribe action under specified circumstances*; and therefore, analogously, we can complain of an axiology if it fails to prescribe action. Thus, for instance, it is certainly a good criticism of an axiology that when conjoined to empirical facts about mind and society it leads to two incompatible action prescriptions

for one and the same person in one and the same set of circumstances and any specified condition of that moral agent's knowledge. (Incidentally, this is one of the most obvious refutations of the Freddy Ayer 1936 view that ethical sentences have no cognitive content.) However, there is another side to that coin that is important here. If two bases for faulting an axiological system are that it (a) *fails to cover a case* or (b) *does so inconsistently*, then perhaps the axiologist ought to get the benefit of that general posture concerning the purpose of axiological systems, namely, that it ought not to be required of an axiological system that it prescribe action for horrible hypos that have never arisen and that are almost certain never to arise. Ethical philosophers are even fonder than lawyers are of bringing up a crazy outlandish "horrible hypo" fact situation that, so far as we know, has never arisen (without a much more enriched embedding context) in the whole history of human moral action, and thereby faulting a particular ethical postulate or system of postulates on the ground that it leads to an intuitively unacceptable result, or on the ground that it does not tell us precisely what to do. Just as a scientific theory cannot be properly faulted for not "explaining" or "subsuming" a fact situation that we have no reason to suppose has arisen or will ever arise, so one might say that an axiological system cannot be faulted for failing to prescribe what is the morally proper action under circumstances which, so far as we know, have never arisen and are not likely ever to arise. In this respect one might immunize an axiology from that kind of hypothetical criticism in the same way that we immunize a decision policy from a kind of criticism that deals with nonexistent decisional situations. Just as in Section 17 *supra* I allow for a certain coarseness or lumpiness or gray region business in ethics comparable to that which we have to put up with in physics; the moral undecidability of a situation "way out in left field" should be allowed for in an ethical system.

There are some interesting analogies between the nomological case and the axiological case which I will simply mention but not pursue in depth here. First, in the nomological case, we may meet with alleged particulars in the observational domain that contradict the postulated theoretical causal system; worry about these corresponds, if I am right in my "positive reconstruction" of ethics, to the requirement (which we all seem to accept even if we have rejected Prichardian piecewise intuitionism) that the axiology ought not to be excessively, or too frequently, or too violently, incompatible with our ethical intuitions *in concreto*. Second, we fault a nomological system if it is inconsistent, which sometimes means that a subset of the postulates required to generate the observations in one factual domain and a subset of postulates required to generate observations in some other non-overlapping observational domain, are mutually incompatible; this corresponds in axiology to faulting an ethical system because of internal inconsistencies or because the system, taken as a total network of prescriptions, and when conjoined with empirical facts about mind and society, generates incompatible prescriptions for conduct. And now third, there are *lacunae*, analogous to what are known in jurisprudence as "gaps in the law" (as contrasted with inconsistencies in the law, or cases where the application of two legal principles *dicto simpliciter* leads to opposite resolutions of the instant case).

The point here is that the requirement of categoricalness or completeness is not met even in the most exact of the physical sciences, although it is a regulative ideal for formal calculi. Nobody in his right mind thinks we have completeness of the postulate sets for biological or social sciences. Why should we insist upon it for ethics? It might be taken as a regulative ideal without taking it as an “instant rationality touchstone” whose mechanical application kills an axiology. “Let us try to increase domain coverage, at least for realizable moral cases,” is a useful *meta-policy*; it should not be put as a tight *meta-rule* of rejection, such as, “If proposed axiology A would not yield a clear and precise answer for a moral agent finding himself in hypothetical and deviant situation S, never yet realized or at all likely to be, then A is refuted; and, being refuted, must be immediately discarded.”

§ 21. I suppose in one sense the “deepest” question of all is one that is, in axiology, rather close to that dreadfully deep question in nomology, to wit, the mind-body tie as the last step in the epistemic chain (which is also, one must presume, the last step in some sort of causal chain). Popper argues that one should not say that observations are “evidence for” protocol statements, because to be *evidence for* something means to be propositional. Only propositions can be logically connected with one another, and a proposition cannot be logically connected with a non-proposition, such as a thing-predicate observational event or a phenomenal state or occurrence. While I am quite willing to go along with that convention concerning what it takes to be evidence for something, at the same time I think Popper should, in consistency with his reliance upon Tarski’s definition of truth, be willing to say that language entries are normally tokenings that accord with a language entry rule, that is, with the semantics of the observational language; but that persons sometimes token in violation of such a semantic rule, such as the degenerate “‘red’ means *red*,” or the more informative (to an English reader who doesn’t know French) “‘rouge’ means *red*.” Such persons then token illicitly and, I should wish to say, “non-correspondently” (given a correspondence theory of truth). Popper says that perceptual occurrences *cause* or *motivate* (rather than justify) the appropriate tokenings (*Logic of Scientific Discovery*, p. 105). I agree, but would add that, while they do not “prove” them (not being propositions), nevertheless they do, in some to-be-carefully-stated sense, *render them legitimate*. That is to say, tokening of a protocol sentence in the absence of a coexistent perceived event that the tokened sentence describes, given the usual (and implicitly relied upon) semantics of the user’s language, is an illegitimate language entry, an illicit move in the language game, “a violation of the applicable semantic rule.” The psychological tie between perceptions and observation-statement tokenings is, I suspect, analogous in a deep sense to the relationship between the feeling of obligation (*or* of guiltiness for not doing what one is obligated to do, *or* the feeling of abhorrence or disapproval toward someone else’s unethical conduct) and the tokening of a Prichardian ethical intuition *in concreto*. Of course, the meaning of words like ‘obligated’ as reports of one’s intuitions *in concreto* is not precisely the same as the use of those words in the reconstructed utopian econometric axiology I have fantasized.

But neither is the meaning of a word like ‘velocity’ as it appears in the embedding text of a highly theoretical network, such as kinetic theory of heat, *precisely* the same as the pre-analytic observational ‘velocity’ of billiard balls and stones which enables us to get our system of meanings going in the first place. Again, I hope without committing a *tu quoque*, I suggest that he who is willing to put up with the current state of metatheory in its reconstruction of complicated empirical sciences instead of deciding to junk the whole business because that reconstruction is not anywhere near satisfactory, and who has decided to reject the upward seepage theory of meaning (nevertheless maintaining the idea of *some sort* of empirical tie to observations and, hence, to our preanalytic “pre-scientific” language of thing predicates) cannot, without contradicting his own meta-principles as to what sort of discourse he will allow as cognitive despite its unsatisfactoriness, reject ethical statements or systems on the ground that we cannot give a complete satisfactory account of the way in which the preanalytic feeling of moral obligation relates to the reconstructed technical word ‘obligated’ as it appears in the generalized abstract axiological system.

In this connection, I mention without elaborating it my impression that even sophisticated persons still have some funny yen to *derive* or *deduce* the distinctively “moral experience” from a system of ethical principles or maxims, which seems to me to be a wholly mistaken demand. One does not expect to derive the quality of the red experience from the language entry rule, “‘Red’ means *red*.” (What on earth would that demand for derivation *mean*, I ask myself?) If somebody were to ask an epistemologist or psychologist to do this, we would simply point out to him that it is an inappropriate (probably nonsensical) demand to make. In the Morris and Co. language of Neurath’s *Encyclopedia*, we would say that the demand arises from mixing two branches of semiotic, i.e., semantics and pragmatics. The language entry rule *qua* rule belongs to semantics, the causal efficacy of a red-hued phenomenal event in eliciting a color-tokening event (and, hence, a long-run statistical “appropriateness” of color talk) belongs to pragmatics. *Mutatis mutandis*, the classification of a contemplated act A_i as morally illicit is axiological semantics; the attendant guilt-feeling (for us non-sociopaths) is a psychological happening we may discuss in “impure pragmatics.” I suggest that there is for an axiology a kind of “language exit rule,” sort of like the semantics of color words but now going in the other direction, in which we would say something like, “‘Obligated’ means *obligated*”; and then there is a *psychological fact* that when you are obligated, you (usually) feel an impulsion (unless you are a sociopath) to do that which you are obligated to do. Furthermore, if you fail to do it, you will then have a feeling that we call ‘guilt,’ and an impulse to make reparation, or to apologize, or to seek and receive some sort of human or divine forgiveness. These dispositions to actions and experiences are, in my view, what correspond in the case of an *axiological language exit rule*, to the color experiences and the associated events that correspond to the semantic *language entry rules* of color observations. The only way in which we could ever “derive” or “deduce” that one will tend to have some kind of guilt experience, or will feel impelled to carry out

some obligatory action, or will tend to feel disapproval of someone's conduct, is a psychological analysis of how people learn to use moral terms and to apply moral discourse to their own conduct and the conduct of others. That job corresponds to the job of psychology of observation and perception and reasoning in the non-ethical sphere. Personally I have never been able to get excited about this one and have never thought it was mysterious, but there may be something that I am overlooking.

§ 22. Part of our residual positivist intuitions against assigning a full-blooded objective reference to ethical predicates is, I suspect, involved with a somewhat simplistic form of the correspondence theory of truth. That is, despite everything we may say in rebuttal of the kinds of familiar objections discussed *supra*, the persistent haunting question remains as to “just *what* and *where* in the external world are the entities (events, dispositions, relations, or whatever) to which my ethical assertions must *correspond* when they are true, and fail to correspond when they are false?” Since the answer to this question is unclear on any current reconstruction, the positivist-correspondentist-naturalist kind of person (such as Feigl or Meehl) is reflexly impelled to commit ethical discourse to the flames. But here again we may be applying a double standard of meta-theoretical morals. The correspondence theory of truth (while perhaps in better overall shape than the coherence theory) has serious problems in science, mathematics, and common life, not only in ethics. Again I want to avoid any whiff of *tu quoque obscurantism*, but I cannot avoid mentioning that the mere reiteration of Tarski's famous 1936 definition of ‘truth’ (which, anyway, was aimed at purely formal systems, was it not?) really does not do as an adequate meta-explanation of precisely what the correspondence theory of truth says for either natural languages or, goodness knows, for scientific languages. I don't think any contemporary logician or philosopher of science (even less any historian of science) would say that we are prepared to give a clean and rigorous account of how we apply the correspondence theory of truth to statements about photons, or virtual displacements in classical mechanics, or the separate chunks of notation that occur in the Schrödinger equation, or even the handy but puzzling notion of centrifugal force. Since we do not cease and desist from serious discussion of nomological meanings and verifications pending a resolution of these terrible difficulties in the reconstruction of theoretical science, it must be that we are willing to continue repairing the raft—the metatheoretical raft—while we are floating around on it. I am afraid we are going to have to settle sooner or later for some strange and subtle fusion of the correspondence and coherence theories of truth in order to give a metatheoretical account of science which is at all adequate to the way things are. Even on the “upward seepage” doctrine of the meaning of theoretical terms, as souped up and articulated by use of the Ramsey Sentence, it was never entirely clear—at least to me—how the “truth” and the “meaning” of theoretical terms thus introduced via the nomological network and anchored at the observation base by a proper subset of those theoretical terms being coordinated (by definition? by bridge law? if by the latter, are they themselves theorems?) to so-called observational terms. But with the growing doubts on all sides as to the adequacy of a

Ramsey Sentence “upward seepage” reconstruction of the meaning of theoretical terms, I think it fair to say that the thing is even less clear than it was twenty years ago when the Ramsey Sentence first began to be widely relied on for this purpose. I doubt that any logician or historian of science is able to give at the present time anything like an adequate account of the meaning and truth value of sentences stated in the theoretical language of natural science, either via old fashioned so-called “operational definitions,” or by the Ramsey Sentence, or by some combination of these with an embedding text involving Mary Hesse’s idea of models, or some combination of all this with an upward seepage theory of meaning. At least I haven’t seen one yet. So it seems to me increasingly plausible to anticipate that the best jerrybuilt sketch of a metatheory that we can have about the truth of theoretical statements in science is some funny mixture of a correspondence and a coherence theory of truth. It is difficult to understand just what a pure correspondence theory of truth means for theoretical entities that are introduced via the nomological network, and certainly it is extremely difficult to do it via the upward seepage theory of meaning, given the universally conceded meta-point that the derivations go downward to the observations (even that is in grave doubt presently) but they never go upward. Hence we get the familiar talk about “partial meaning specifications” or “incomplete contextual definitions” and the like. Now if the very *meaning* of the theoretical terms that supposedly “correspond to” certain alleged entities in the external world is incompletely given by the network, it is kind of hard to understand in what sense the correspondence of sentences about those terms to the non-verbal reality can be itself a clearly defined logical (semantic) notion, is it not? Again, let me emphasize a point that I have made in several different connections during this memo, to wit, that the intellectually responsible metatheorist ought to get as clear as he can about his *meta*-meta-criteria before he begins applying meta-criteria to a subject matter domain, especially when the intent of such application is to liquidate concepts. I don’t see why we feel that we have to have a pure correspondence theory of truth as applied to ethics, and a completely satisfactory one, when we haven’t managed to do it for science.

§ 23. Without surreptitiously trying to avoid the task of giving an affirmative meta-theoretical reconstruction of the content and “objective correspondence” interpretation of ethical sentences, I think we can rebut two common transitions made by persons of my generation raised in the logical positivist tradition, each of which moves too easily to the conclusion that ethical sentences lack truth value. The rebuttal need not consist of offering an affirmative metatheoretical reconstruction, but consists in challenging the positivist’s transitions. Maybe these transitions are somehow possible, but the point is that the logical positivists have usually made the transitions *instanter*; and if one states a rule or principle in the metalanguage that would be capable of making the transitions *instanter*, the rule that seems to be required does not look in the least plausible, and is surely not analytic.

First Transition Premise: “Granted that cognitive ethical discourse can occur about validity of derivation chains, or coherence with facts, or means-end relations, or accord-

ance with our moral intuitions *in concreto*, it still remains true that the primary ethical statements, the distinctively axiological premises of such a system, are pure postulates. That is, they function as the premises of arguments, not as the conclusions. They are only valid in the logician's trivial sense that the postulates of a system are always taken to be valid." From this, which will be admitted, there occurs in positivist thinking an immediate transition to the conclusion: "*Therefore*, while the theorems may be 'valid,' these postulates cannot be 'true' or 'false.'"

Now, as I say, it may be possible somehow to get from the metatheoretical premise to the conclusion, but the "therefore" is obviously not justified without some unpacking. Such a conclusion would not be drawn in connection with arithmetic or pure geometry or logic or set theory; nor would it be drawn with respect to any empirical science. Every argument begins with premises; every system of propositions related definitionally and deductively to one another (whether categorical or not), involves reliance on a proper subset of propositions which are not themselves derived but which are used to derive the others. If there were a metalinguistic principle to the effect that, roughly, "Any proposition which appears in a given system of propositions (or, in a given string of discourse) not in the rôle of a conclusion but in the rôle of a premise or postulate or axiom cannot have a truth value, cannot correspond to anything that 'exists' or 'is valid' or 'is correct'," it would generate troubles (or solve them high-handedly) in all of the formal and empirical sciences. The questions whether and in what sense the postulates of, say, arithmetic are objectively true or correct (in addition to being trivially valid since they occur as the premises of the formal system in question) are, as I understand it, among the profoundest questions that exist in meta-mathematics. I do not believe there is any consensus on the matter, and my impression is that many mathematicians today would want to say that they are, *in some sense*, true or correct. Take the fifth of the Peano axioms. It is a postulate, which cannot be proved as a theorem without circularity, and upon which the whole of arithmetic (and hence all of mathematics) depends. So the Peano Fifth Axiom, the Principle of Mathematical Induction or however you want to label it, is a "pure postulate," in the only sense in which I can understand that phrase (since Williams did not define it). But do we want to say that it has no correctness or (non-trivial) validity or truth-value, that it does not correspond to "the way things are" in the realm of numbers? I don't believe there are many mathematicians who would want to say this, and I don't think many logicians would either. Perhaps Carnap would have, if pressed; but I rather think even he would be uncomfortable about it. I would be willing, as a non-expert in this area, to go out on a limb and say that Peano's Fifth Axiom, *in addition* to being a "pure postulate," in addition to playing a critical rôle in all derivations of arithmetic (and being presupposed for the interpretation of all of them) is also true or correct. That is, I sort of believe it is actually true in the realm of numbers (wherever and whatever that realm is, in Plato's or Kronecker's Heaven!) that if the first member of a sequence has a property, and if the property is one which, if it belongs to an element in a sequence, belongs to its successor, then it *is the case* that all of the elements in the

sequence have the property. It seems to me that's near as "self evident a truth" as the Law of Non-Contradiction or the Rule of Detachment. Or, now that I've mentioned it, take the Rule of Detachment: We usually state it as a rule, and not as an assertion. I don't quarrel with that practice. But doesn't everybody believe that it is, in some fundamental sense, "correct" or "true"—that even when conceived as a rule, it is a proper, unavoidable, necessary rule? It's not only that you can't do any kind of logic without the Rule of Detachment, and that even somebody who wants to argue about it (I haven't ever met anybody who did) would find himself presupposing it or relying on it, otherwise his argumentation could never get off the ground. It's more than that, don't you think? Surely we all believe, even if we would try to avoid saying it in the interest of some kind of logician purity, that the Rule of Detachment is a *good rule*, because it says *the way things are*. That is, if we permit ourselves the "material mode of speech" (and why not?) we would say, "It's true that you need the Rule of Detachment functioning *as a rule* to carry out argument and to make derivations in logic, set theory, etc. But we also believe that it is in some basic sense "correct" as a statement of the transmissibility and preservation of truth. We believe that if somebody decides not to employ the Rule of Detachment he is not merely estopped from discursive thought or rational argument; he is failing to embody in his cognitive practices a deep metaphysical truth about the way things are, because (we metaphysically believe) if you are safe in asserting a proposition, and also safe in asserting that another proposition follows from it, then you are "equally safe" in asserting that consequent proposition." Despite Carnap and Co., I have never been able to see this as purely a "useful convention" or "arbitrary stipulation" about how a formalism is manipulated, have you?

It is needless to multiply examples. Point: We often attribute truth value to statements that are postulates or premises or presuppositions; and there is, to my knowledge, no metatheoretical law saying that anything that is a postulate rather than a theorem, a premise rather than a conclusion, or anything which we presuppose without explicitly premising but can be made to realize that we have presupposed in the course of our argument—that no statements of this sort are true or false, or could have any correspondence with reality. It seems to me that one would be high-handedly settling some deep issues in meta-mathematics, and he would be adopting a dogmatic formalist position over against a logicist or intuitionist position, if he said, "If p is a premise, it cannot have a truth value." Furthermore, he would be construing the formalist position as a very strong "reductionist, nothing-but" position (which I don't see how you could possibly get *out* of any formalism, could you?) to the effect that not only is formalism a perfectly adequate account of everything that is done in meta-mathematics and logic, but one can be quite certain, since it is an adequate account (which has not been shown to the satisfaction of the opposition, has it?) that truth value in formal systems amounts to formal validity or derivability or deducibility *and to nothing else*, and therefore the only truth value that the postulates could have in any kind of formal system is the unexciting logician's meaning of "valid" as applied to postulates rather than theorems. All this appears dubious, and if applied in

meta-mathematics would be viewed as high-handed, dogmatic and exclusionary; yet I suggest to you that we positivist types are strongly inclined to do it in ethics when we are confronted with somebody's saying that the primary axiologicals of an ethical system are "pure postulates." We immediately infer, as if it were a direct consequence of that statement, that such postulates *could not be either true or false*. I repeat that there may be some *other* kind of meta-consideration (especially unanswered criticisms of all available affirmative reconstructions of the truth value or correspondence of ethical sentences with whatever "objective" they are supposed to match or fit) that would lead to this. My point is only that you can't get it *instanter* from the meta-premise, "Ethical axioms are pure postulates."

Second Transition Premise: "When we say a primary axiological proposition is a pure postulate, we mean that it functions as a premise in all ethical argument but it is not induced from empirical observations nor derived as an axiological theorem within the system. But since ethics is about action and embodies evaluations, one non-technical way of saying that primary ethical propositions are pure postulates is to say that they express not observations and not conclusions or deductions or causal explanations or existence claims, but rather they express the individual moral agent's *basic commitments*." Conclusion: "Therefore, anybody's commitments are as good as anybody else's; mine are as good as yours, Hitler's are as good as Buddha's, and so on." Here again, we know that a case can be made for a radical ethical relativism (despite the anthropologist's insistence that there is a considerable common core running through all cultures), and an affirmative account of ethical intersubjectivity should be provided in ethical metatheory just as an affirmative account (in causal as well as logical terms!) should be provided in theory of (descriptive) knowledge and in philosophy of science. Again, I am not claiming to rebut a whole battery of unspecified arguments for radical ethical relativism. I merely point out that the transition *instanter* from the above premise to the conclusion is dubious, to say the least. We wouldn't hold still for any such thing as that in any other field than ethics, would we? In non-axiological domains, we do not operate with a strong meta-principle to the effect that, roughly, "Since cultures or individuals differ as to their rock-bottom beliefs (concepts, paradigms, metaphysical commitments, ontologies, ideas of lawfulness, criteria of truth or explanation) *therefore* one culture's or one person's rock-bottom beliefs or commitments must be as good (sound, true, correct, valid, useful) as another's." It is notorious that one cannot argue a primitive man out of his belief, say, that the sun is god by any amount of appeal to what we Western Faustian technological minds would call "scientific reasoning." So I suppose one has to say that ultimately the basic propositions are commitments, whether they are basic in Popper's or Russell's (epistemic) sense, or basic in the other (ontological) sense that they are the theoretical premises about the way the world is, by which observational statements are interpreted and their relevance assessed. But nobody concludes from this that the metaphysical beliefs of a tribal society about the sun are just as good as Feigl's belief that the sun is a fiery ball of mostly hydrogen. (Maybe a few anthropologists conclude this, but you and I

would see them as soft-heads, would we not? But that's a diversion.) The positivist argument as it stands says that if one acceptable gloss on the statement "Primary axiologicals are pure postulates" is "Primary axiologicals serve to express the basic commitments of the person who holds them," then the immediate transition can be made to, "Therefore, what any two people hold as to their basic commitments cannot be set in opposition to one another, cannot really contradict one another; and it cannot be said that one is preferable or better or truer than the other." Again, I say that there is obviously not a tight logical transition here. It must be supplemented by some other general meta-premise, which would have to say something like the following: "Whenever two individuals have differing basic commitments which are not derived by them from facts or postulates, neither can in any sense be preferred as better or truer." I think D. H. Monro has answered this (see his excellent *Empiricism and Ethics*, 1967) but my answer would be simply that I don't see any reason why I should buy that meta-premise. Do you?

I don't have the Monro book available, but he pushes fairly strongly (and I think convincingly) against that form of relativism or "ethical arbitrariness" which says that if I affirm a value or express a commitment (and call it that) I am automatically bound to say, by implication from my so describing it, that your commitments are just as good as mine. If you contemplate that one for a while it gets to look pretty strange, does it not? "I am committed to avoiding cruelty" is one way to state a primary axiological (at least I would so see it, although Sartorius would presumably try to derive it); but the language about "commitment" may be kept entirely metalinguistic. I could instead say, in the ethical object language, "Cruelty is wrong." (I rather suspect this is one that does not have to be qualified by a statement of *prima facie* wrongness or defeasibility, whereas if one said, "Producing pain in a sentient being is wrong," it would have to be so qualified, as when the pediatrician has to hurt a child in curing an illness.) Or, using the distinctively ethical term which I think is probably the only unavoidable and non-reducible ethical primitive, the word 'ought,' I can put it, "One ought not act cruelly," where "act cruelly" means "induce suffering in a sentient being against its will and not as a means to helping, healing or otherwise benefiting," in other words, producing pain as an end in itself, making somebody hurt just because you like other people or animals to hurt. I think that cruelty is wrong, as I dare say you would agree (if anything is wrong). If I state this in the ethical object language, and then you ask me what its status is, I'd tell you (in the metalanguage), "It's a pure postulate." And then I might have, as a gloss on *that* metalinguistic statement about pure postulation, the further comment in pragmatics: "It, like other pure postulates of ethics, expresses my basic commitments." As I believe Monro first stated clearly enough so I got the full punch, it is unnecessary for me to affirm equal status for counter-commitments when I express in the object language a basic moral commitment like "Cruelty is to be avoided" or "One ought not to act cruelly" or "To do cruelty is wrong," or, even more personally, "I abhor cruelty," or even in today's familiar imperative or hortatory mode (which I would not accept as a translation of axiological discourse, but many do) "Here you, avoid cruelty!", or even the limbo

status “Cruelty, ugh!”, or the optative “Would that neither I nor anyone else were ever cruel.” None of these forms as it stands clearly entails “If someone other than I, the present speaker, were to utter ‘Hurrah for cruelty,’ then I would be obliged in consistency to say, ‘Hurrah for his views.’” In fact, if any argument at all can be made via a reasonable meta-language premise on this matter, it would seem to be the other way around, as I believe Monro argues. That is, “Would that everyone would abhor cruelty, as I do,” if it implies anything at all about other people’s statements about cruelty and abhorrence, would seem rather to imply “Would that everybody, when asked to express his basic moral commitments, would say, as I now say further, ‘Would that everybody abhorred cruelty as I do.’” That is, maybe I can’t get from my optative or hortatory or imperative or primary obligation statements to *any* meta-statement I might make about other people’s primary or basic ethical statements. But if I can get anywhere, via some non-arbitrary and other than ridiculous meta-principle, it hardly seems that I could get to the statement that I approve of your approving things I disapprove of! If *any* such derivation is possible, which it may or may not be, I should suppose that what I could derive is that I ought to disapprove of your failure to approve what I approve. In summary: I can’t generate “We are all equally right when we disagree about ethics” from the meta-linguistic statement that the object language statements I make are primary postulates or express my basic commitments. Nor could I derive this about you from your value-commitments, or from your meta-comment that “commitments is what they are.”

I don’t deny, as I have said several times in this memo, that there is a terrible problem about the objective (and not completely intersubjective) status of ethical statements. But I think this particular way of looking at it is misplaced in emphasis. It seems to me that the two really serious stomach aches involved are (1) the epistemological stomach ache, that is, of setting forth a respectable metatheory sketch of how ethical statements are confirmed (as a neo-Popperian I would rather say how they are tested, since I don’t see why we have to be justificationists about ethics when we are not justificationists about science); and then (2) a satisfactory metatheory sketch of the ontological status of the ethical relations spoken about (if we are to hold in some degree to the mixed correspondence-coherence rather than the pure coherence theory of truth). I think that the combination of the epistemological and ontological troubles can be expressed in a variety of ways, but I think this particular one is misleading and unfortunate. It tries to place at a very elementary level of meta-comment on primary ethical statements as pure postulates a difficulty which is rather more complicated. In trying to locate it there, it commits a plain blunder, because it attempts to go from the meta-comment that somebody’s basic axiological statements are functioning logically as pure postulates (rather than as observation statements or formal theorems), plus a gloss on that statement amounting to some form of explaining why they are pure postulates insofar as what they express is a person’s primary or rock-bottom ethical commitments, to the further meta-statement that one who holds that view of them must, in consistency, hold that other people’s primary

commitments are as “good” (valid, true, correct, or whatever) as his are. Even if that latter one could somehow be shown, it simply does not follow from the premise.

General comment on these common positivist transitions:

There is an interesting analogy, vague and imperfect but nevertheless illuminating, between this problem and the problem of premises in empirical science. The premises in empirical science that play the rôle that primary axiologicals play in ethics are, of course, the system of postulates about theoretical entities which, on the old view, got their sole meaning by “upward seepage” from the observable facts. But in terms of what used to be called inductive logic, the “premises” of the scientist’s argument are, of course, the protocols (observation sentences, or what Popper calls basic statements). There is a legitimate interpretation of each of these as premises. In examining the elaborated structure of a good nomological net in a given field of developed science we would usually take the postulates as premises when we are working our way through a theoretical derivation of experimental results with an eye to “understanding how the theory works.” But if we are talking philosophy of science with emphasis on epistemology we would usually take the protocols as the premises of the “inductive” argument. I don’t see that there’s anything to dispute about there, it just depends upon whether you are tracing out the derivation chain of facts from a theory, in which case the theoretical primitive statements are the premises of the arguments; or whether you are instead trying to make a case in favor of the theory “on the basis of the facts known from observation,” in which case the latter play the rôle of premises. I want to focus on the latter case for a moment. Without trying to mediate between Nagel, Carnap, Kordig, Feigl and Co. on the one hand versus Popper, Feyerabend, Kuhn, Lakatos and Co. on the other it’s very important to realize how tough that controversy can be to resolve. Why should we be so anal in our meta-criteria for satisfactory resolution of ethical problems, given the sorry state of such a basic problem as this in philosophy of the physical sciences? I think it can be safely agreed that Popper is *in some sense* right when he says that including a basic statement in the corpus can be causally explained or psychologically motivated but not “logically justified.” And I am amused at how rarely Kuhn, Feyerabend, Polanyi and Co. even mention Neurath, who, if I recall correctly, traumatized some of you fellow positivists way back in the early 1930s by saying explicitly that whether we include a proffered protocol into the corpus is a *decision*, and hinges very heavily (he was surely right about this for single protocols—the trouble is that he gave no account of their aggregate control of the network) upon whether we can fit it in (*eingliedern*). This is the position that led Lord Russell (in his 1940 *Inquiry Into Meaning and Truth*) to comment that it seemed strange to him that a philosophy which began in the spirit of empiricism could end up saying, with the author of the Fourth Gospel, “In the beginning was the Word, and the Word was God.” Nobody has as yet reconstructed precisely how the mass of alleged (purported, “candidate”) protocols in the aggregate control the nomological network despite the fact that the nomological network usually (not quite always—when not?) controls the admission of an individual proffered protocol into the corpus. But it is

blindingly apparent (to me at least) that *something like this* is what happens, and some reconstruction of it must someday be given. That makes me a bit happier about ethical discourse, because the same seems to me rather clearly true about ethical reasoning, where it gives us undue discomfort. We revise some of our gut level Prichardian intuitions about ethical cases *in concreto* by reflecting upon our axiological network; but we also require of our axiological network that it should be, “by and large,” in some generic or collective sense, in accord with the mass of our reflective ethical intuitions *in concreto*. I conjecture that the logical problem here is exactly the same as in Neurath’s Problem, despite the great differences in what one can appeal to as argument and evidence and the obvious differences in the domain of entities spoken about.

Anyhow, I digress. What I mean to emphasize about the two routine positivist moves is that when we take the basic statements of common life or of physical science, and (goodness knows!) of social science as premises, we are forced to recognize that not everybody on the planet Earth shares them. Now you as an old positivist may give me here the standard Vienna answer, which I don’t reject as without merit but which I think tends to come rather too easily to persons like you and me, “Well, of course, we don’t expect a tribal society to make very good scientific observations, because they don’t know enough.” And associated with that, often unexpressed but easy to smoke out from us old positivist types, is the negative statement, “On the other hand, a comparable assertion about an unenlightened tribal society *cannot* be fairly made about their ethics.” But why not? I don’t think we want to make a rock-bottom qualitative distinction here on the basis of the *degree* to which you can twist the arm of a tribal society to accept the technology of the machine gun for fighting their wars with enemy tribes, versus, say, convincing them that Freud’s theory of dreams has something going for it, versus, say, convincing them that the universe started with the Big Bang rather than being hatched from an ostrich. These are matters of degree which we are not in the habit of allowing to be distinguished in the positivist tradition in a qualitative way. We have customarily declined to draw sharp qualitative distinctions between, say, psychoanalysis and Skinner’s learning theory and molecular biology and Keynesian fiscal theory and the theory of the chemical bond and quantum mechanics, right? It was always part of the positivist line to insist, for instance, that Dilthey’s distinction between *Geisteswissenschaften* and *Naturwissenschaften*, while it pointed to some important differences of degree and to some extent also in methods of investigation (even including, for some positivists, the greater rôle of the investigator’s own self awareness, at least in the context of discovery, for the *Geisteswissenschaften*), never meant fundamental differences in kind. But I needn’t belabor this point, since I realize that you would agree with me about it. We do not have to buy all of Feyerabend, which as you know I do not, or Kuhn’s touchiness about the concept of truth, in order to agree that complete unanimity is lacking among verbalizing bipeds (I won’t say “rational persons,” because somebody might get to nitpicking about how rational the primitive mind is) as to what basic statements are admitted. Instead we may talk about the “premises” not as the basic (protocol) statements

required in reconstructing the testing of scientific nomologicals, but as the theoretical premises of the nomological network, or, if you like, the basic kinds of things which the world is supposed to be made up of (see the fascinating book by Claude Levi-Strauss *The Savage Mind*, especially chapters “The logic of totemic classifications” and “Categories, elements, species, numbers”). One knows that a tribal society, medieval Catholic man, and modern secular scientized man differ strikingly and pervasively in how they carve up the world of things and concepts. I am with Feigl, Kordig, and Carnap against Feyerabend, Kuhn, and Co. (although I don’t want to do Kuhn an injustice—maybe I should just say Feyerabend) that it *ought* to be possible with sufficient care, and that in general it *is* possible with sufficient care, to get anybody who is not feeble minded or psychotic, and who has corrected 20/20 vision, to read an ammeter correctly. And next time I see Feyerabend I’m going to really push him on Campbell’s 1919 book, where Campbell discusses the fact that, and the reasons why, 99.9% agreement among sane persons of normal perceptual equipment can be reached in respect to three kinds of “raw observational data,” to wit, temporal simultaneity or betweenness, spatial coincidence and betweenness, and judgments of number. I repeat, I side with you on this question, partly because I see some psychologists as tending to use Polanyi’s views for obscurantist purposes. Nevertheless, you and I do have to admit that merely confronting a high IQ, normal vision and hearing, non-schizophrenic chieftain of a tribal society with the set of observation statements that Millikan made in his oil drop experiment just won’t do, it won’t convert him. Everybody knows this. Now what we positivists were inclined to say—I think now with undue ease and rapidity and not worrying enough about what was packed into it—was, “Well, of course, you have to educate him to make proper use of instruments, and to make the proper correction for his mistaken beliefs about the dream world, and to recognize that there are such things as hallucinations, so that when he sees a ghost after eating peyote or fasting before going out on his adolescent solitary hunt he shouldn’t take it seriously as a protocol sentence describing external objective facts, etc., etc., ...” Now, this is all true, but one has to realize how much he is saying when he says this. He is saying that you can’t get this member of a tribal society, despite his normal mentality and sensory equipment, to see the right things in the right way and to dump the wrong things into the observational trash heap because he doesn’t understand—what? Just what is it he doesn’t understand but we do? *The received network of Faustian scientific man*, that’s what he doesn’t understand. God forbid that I should say that the received network of Faustian scientific man is no better than the received network of tribal societies; you know me better than that! My view is very close to the old fashioned Victorian physicist’s view that tribal societies have a mistaken picture of the world; or, better, that we have one that is only somewhat mistaken, whereas theirs is very largely mistaken. I am *not* talking some bobby-soxer cultural relativism about physics; in fact I would not even hold cultural relativism for psychology, as you know. The question here is not whether the two networks differ in their verisimilitude, because we started out talking about argumentation and epistemology, not about ontology. Whatever may be the

verisimilitude of a tribal society's nomological system versus Western scientized man's nomological system, it remains the case that there is sometimes persisting disagreement about what some would think of as basic statements in inductive inference; and clearly there is even more persisting disagreement about premises starting in the other direction, that is, theoretical entities explaining the observed, even when we can get agreement on the observed; and when we say "in principle" you can bring the member of a tribal society around, it turns out that what that really means is that you can *sometimes* (note, we have to admit with the anthropologists, *not always*) bring a sufficiently massively exposed and tractable member of such a society around to entering our nomological network and dealing with the protocols accordingly. And with that I come back to my core thesis in this memorandum, to wit, that one cannot now say intelligently what the situation is in axiology because the state of axiology as a cognitive discipline is as primitive as Hume's picture of empirical science was. So that when the positivist says that "in principle" you can get members of the tribal society to agree with us about the nature of the sun, he doesn't *really mean* you *can* convert *all* of them. (I believe the anthropologists will tell you that you can't convert most of them.) What he really means is *if* you can get one of them to buy *enough* of our network, then it will be *fairly* easy to get him to buy *most* of the *rest* of our network; and he further premises that if this person "saw the world rightly" he *would* buy it. And that's quite a different claim from saying that *in fact*, if he's non-psychotic and not feeble minded and has normal vision and hearing, he *will* unfailingly be made to buy it when we give him a couple of lectures on classical mechanics or Freud's theory of libido or the germ theory of disease. I claim that when you read *Mein Kampf*, which I had occasion to re-read a couple of years ago, you realize that Hitler's axiological system, quite apart from our historical knowledge of what it led to in liquidating six million Jews, but just *taken as it stands* without knowing what followed a decade after he wrote it, can quite fairly be described as a mishmash of miserable crap, creaking at the seams, inarticulate and contradictory and in some sense arbitrary and capricious in the extreme. That is, "Hitler's ethic" is about as jerrybuilt a conceptual mess, reflecting the twisted wanderings of a third-rate uneducated mind, as the Azande conception of witchcraft is revealed to be by such a sympathetic writer as Evans-Pritchard. You can see what I am getting at here. I suggest that there is an important resemblance, although great differences in degree, between the cognitive situation of tribal societies, who can hardly assimilate Western science in its conception of astronomy and of the causes of mental disease, and *most of us today struggling with complex ethical situations*, where the equivalent of Western science would be the kind of utopian econometric-ethical formalism and postulates that I have fantasized throughout this memo. The connection this has with the two easy positivist transitions is hard to state, and I may have lost the train of thought a bit here, but what I'm mainly arguing is that the positivist's reliance upon the existence of different group and individual ethical belief systems is excessive because it implicitly compares this "empirical ethical relativism" unfavorably with an alleged *non*-relativism about science; whereas the latter does not in

fact exist either historically or among different culture groups at the present time. And I want further to argue that the usual positivist rejoinder to this rebuttal, which is to say that if the tribal society members would only “think straight” and “observe accurately” and give themselves some time to learn differential calculus and molecular biology and so on, they *would* understand why we don’t accept a certain observation in the rain dance as germane to the question whether the sun is God or not, that this counterfactual is an unfair rebuttal to the moralist because it isn’t actually a statement about what all persons of normal intelligence and sanity and vision *can in fact be induced to hold*. It is a contrary to fact claim about what they *would* hold if they *were* rational in a special and highly sophisticated sense—a sense that already includes the acceptance of most of Western scientific theory. I repeat once more that I am *not* saying that Western post-Galileo scientific theory is no better than the world view of tribal societies, which is not the point of this discussion (and which I thoroughly disbelieve). The point is that in discussing the “non-coercive” epistemology of ethical sentences one can turn around to the positivist and say that the same is true of basic statements and of theoretical premises outside ethics. When I utter an expression of a Prichardian intuition in my ethics, I do not commit myself to the statement that Hitler’s different intuitions about mass murder must be “equally valid as mine” just because he has hard-to-influence commitments and so do I.

I’m afraid I may have screwed up the line of thought in the preceding paragraphs by paying so much attention to the tribal societies case and the Hitler case, but I do believe they’re deeply relevant and similar in respect to the problem of “inter-subjectivity” and, because of that, in respect to the two meta-inferences I am challenging here. Maybe I can summarize the reasoning more concisely, now that I’ve got tribal societies and Hitler off my chest. A positivist critic of the cognitive status of ethical statements relies in sort of a joint or double-barreled way upon the metalinguistic admission by an ethicist that, while ethical systems as a whole can be tested against facts (because they can generate contradictory moral instructions, given the empirical order) and can be examined in their applications by empirical methods (especially means-end questions and empirical questions about people’s actual sentiments and preferences in utility theory), and while ethical derivation chains can be challenged for validity, and all the rest—nevertheless, it still remains true that, as the positivist insists and as the ethicist (cheerfully or maybe not so cheerfully) concedes, the “axiological primitive statements” are pure postulates. This metatheoretical analysis of the logic of the situation is, I repeat, not very earthshaking, since *any* system of beliefs has some premises and some conclusions; furthermore, in some systems of beliefs, such as the empirical sciences, it is not arbitrary which is which, although, as I understand it, in purely formal systems there is always (almost always?) considerable freedom as to which statements we take to be primitives and which we derive as theorems. The positivist argues from the fact that in a given axiological system there are primitive statements such as “cruelty is wrong” or “one has a *prima facie* obligation to keep his promises” that are not derived as theorems and not deducible from empirical fact statements (although, as I said elsewhere in the memo, they are in some

funny sense testable against “fact statements” of the kind that express our ethical reactions *in concreto*) that they have no truth status or correctness, that they do not “correspond to anything objective.” I repeat that this may be the case, and it may be possible to argue persuasively for it; but it cannot be argued on *this* particular basis. And if the positivist unpacks it, as he can usually manage to do with what I admit is considerable plausibility and reasonableness, by saying that it isn’t just the fact that premises are premises and not conclusions, or that in any string of discourse some things are assumed before other things can be proved or rendered believable or plausible relying on the first assumed things—no, it is more than that, it is that you can push people long enough and hard enough so they will, unless they are “crazy” or otherwise aberrated (an element of circularity here, but not a vitiating one), they will “come around” to sharing most, if not all, of the required basics. Whereas it is alleged that this is not true in axiology. And to this I make a double-barreled response, one factual in the sense of present facts of human conversations about the world, and the other also factual but predictive, that is, concerning the future state of econometricized, formalized, axiomatized and numerified, utopian ethics. The first one, about present and past historical and social facts, is a flat denial of what the positivist says if it is taken *simpliciter* as he states it. That is, it is simply not empirically true that you can *always* get *everyone* who is not psychotic or blind or has an IQ of 50, to agree with what you say about nomologicals or even about observations. It is simply not true that if we screen a particular member of a tribal society for intelligence, sanity, normal vision and hearing, etc., then given a hundred hours of careful instruction he will come to agree with us about molecular biology, the non-divinity of the sun, the causes of pregnancy, the absence of true witchcraft, and so forth.

Nor in our own Western scientific society can we *always* convince *everybody*. There are still psychiatrists who believe that schizophrenia is caused by a battle-axe mother and that it has nothing to do with genes. I happen to know about a psychiatrist in New York who engages in what I would myself consider malpractice and who, I will bet you at least even money, could be successfully sued in a malpractice suit, who has such a passionate disbelief in the genetics of schizophrenia that, if he has a female patient who wants artificial insemination because her husband is sterile, makes it a regular practice to obtain semen from schizophrenic male donors, to show that he doesn’t believe there’s anything genetic about the disease. We know from the Rosenthal data and from Heston’s data that about 1 in 8 of the children thus produced will develop diagnosable clinical schizophrenia, and around half will have recognizable mental disorder. I think that the mother of such a schizophrenic child would have a damn good chance to collect damages from this man. But do we have to assume that he himself is diagnosably insane? Or blind? Or deaf? Or illiterate? No, he is simply a rabid environmentalist fanatic, like the psychologist Kamin who has recently written a book (my old co-author Lee Cronbach says it is so full of logical and statistical fallacies that it’s not worth reading, so I haven’t read it, but I have read the book-jacket summary of his conclusions) which says there is *nothing*

genetic, even within race and class, about IQ. Now this amount of misreading of the mass of evidence we have collected over the last 70 years on the subject may seem to you or me to be almost psychotic; but people tell me that when you talk with Kamin you realize that he is an environmentalist fanatic but that he is not stupid or insane. The point was made once by John Dewey, whom I usually don't quote in matters epistemological (but I see more merit in his views than I used to), in his famous story about the fledgling psychiatric resident who was trying to convince a paranoid schizophrenic patient who had the delusion that he was dead by pricking the patient's skin so that the blood flowed, whereupon the patient said blandly, "Aha, so, now we know that dead men sometimes bleed." I guess everybody admits the point of Dewey's anecdote even if he doesn't like Dewey's general theory of knowledge. Nothing—read, *nothing*—can shake a belief system held passionately enough, whether the Duhemian thesis in its strict form turns out to be correct or friend Grünbaum is right against Duhem. And of course I have always agreed with you that when we test certain subregions of the network we tacitly or explicitly presuppose others, and furthermore it is not completely arbitrary that we presuppose others because usually (not always) the ones we take for granted are the ones that are better corroborated and the ones we are testing have usually at the stage of our knowledge a lesser degree of corroboration. The fact there are psychologists and psychiatrists and sociologists who continue to believe that genes have nothing to do with human traits of the behavioral kind and, in particular, that they have nothing to do with the proneness to mental disorder or individual differences in intellect, is so far out that it is hard for me not to think of these people as a little nutty. But I had a conversation a year ago with Anne Anastasi, who is not psychotic or anywhere near it (and who was recently elected President of the American Psychological Association!) and whose book on differential psychology is one of the three standard textbooks on the subject, who looked me right in the eye and said, "I don't believe that *any desirable traits* are inherited, only defects like color blindness or Mendelizing mental deficiency." What do you say about an intelligent woman with a Ph.D. in psychology who has written a textbook in differential psychology and achieved that degree of professional eminence, who talks such incredible nonsense in the face of the evidence? What you have to say is that she is a rabid Eastern type "sloblib" environmentalist psychologist who is not about to admit that some people are born smarter than others, come hell or high water. And since there is no such thing as hammer-blow deductive proof in the empirical sciences (except in the interesting kinds of cases described by Jon Dorling, which involve a major premise that somebody already accepts and furthermore the acceptance of a set of auxiliary hypothesis for testing), anybody in the inexact sciences who makes up his mind that he *won't believe* that heredity has anything to do with genius or musical ability or social dominance or schizophrenia will always be able to do so. By some combination of *ad hoc*-ing, and setting differential standards of proof for portions of the network, and refusing to admit certain kinds of evidence because of imperfect accuracy and the like, it will be possible, *in my opinion indefinitely*, for ideologues to maintain the dogmatic environmentalist

position. Now here you may wish to say that if they read *and understood* the recent paper by Imre's student Peter Urbach on progress and degeneration of the IQ debate (*British Journal for the Philosophy of Science*, vol. 25(2 & 3), 1974) they *would* see the error of their methodological ways. But note that we positivists, if we argued that way, would again be talking not in descriptive pragmatics but rather in something between pure pragmatics and "methodology," because we would be talking normatively, just as we would say that if somebody had been brought up in the right social class he would know how to speak correct grammar even though in fact we don't seem to be able to get him to do it. I don't think anybody is going to cure Anastasi of thinking as she does about IQ. But she is not crazy, and she is not stupid. So that statements in the pragmatic meta-language about what people counterfactually *would* agree on *if they were more rational* than they in fact are cannot really be statements in the sociology of knowledge or the psychology of belief, relied on for any kind of epistemological argument about "intersubjectivity." They are actually statements of a *normative* kind concerning what the "ideal reasonable mind" *would* do under such and such circumstances. They are not purely descriptive, they have prescriptive components. Point: When the positivist says that you can get well-nigh universal agreement among sane persons of normal sensory equipment, that is literally a false statement, both about the past and about the present. The fact of the matter is that most people's heads are full of [thoughtless nonsense] about most subjects of even moderate complexity, as reading any Gallup poll will show. The universal intersubjective agreement talked about here is an agreement of normatively defined ideal intellects, of the sort that the "perfect natural scientist" is supposed to exemplify (but which, as Feyerabend and Kuhn point out, he rarely does).

And my second factual reply refers to the future. What I want to argue is that progress does take place in ethical, juridical, and political thought, not as fast and not as rigorously as in the case of physical scientific thought but, I maintain, at least as fast as takes place in most thought in the social sciences. Even Milton Friedman or William F. Buckley do not find it comfortable to say, for instance, that a 70 year old man should be allowed by the society to eat out of garbage cans, sleep under bridges, go without medical treatment, and die of exposure in the winter because he was apparently inept or improvident and so did not provide for his old age, he gets just what he deserves. Two hundred years ago many intelligent and cultured people could be found who argued precisely that. In the 1800's Anglican bishops and members of Parliament could be found who agreed with factory owners that it would be wicked to restrict a ten year old's working day to 10 hours of toil, because the excess of leisure time resulting would tempt to vice and indolence. It is true that the Nazis and the Bolsheviks returned to torture as a standard means of criminal interrogation (although, be it noted, here only for "political" offenses, not for ordinary crimes as was done in the 14th century); but it is also true that they felt obliged to keep it a secret and that the rest of the world viewed it with horror. I maintain there has been a growing consensus among rational and ethically sensitive persons in the last three centuries on matters of this kind. It is not necessary for my

purposes here to argue that convergence takes place just as rapidly or as near-universally about ethics as it does about physics. I am quite willing to concede that ethics will be toward the extreme nonconvergent end of a distribution of subject matters (esthetics and theology still farther out). But, alas, I have to remember as a psychologist that the social sciences will also be closer to that end of the continuum than physics or organic chemistry. I don't see that any big issue hangs upon this matter of degree, especially since I am permitting myself throughout to entertain the fantasy that an axiomatized and numerified formalized econometricized axiology of the future will show much more clearly where incoherencies exist and will inform the ethical thinking of cultured minds just as statistics and a sophisticated idea of causation informs the thinking of cultured minds about the physical and social order today. The important point is that you can't get anywhere about the ontology of any kind of sentence by pointing out that people who are not otherwise classifiable as mentally deficient or crazy can persist in being flat earthers or 100 percent environmentalists about schizophrenia. At most, it's a reflection of the state of the subject matter plus the fallibility and tendentiousness of the human mind. It is not a proof of the impossibility of increased knowledge or ontological validity.

§ 24. Let us conjecture that axiological rules, principles, and policies (which for convenience I shall refer to generically as "axiologicals" by analogy with the corresponding scientific usage "nomologicals") are related to concrete ethical intuitions much as descriptive generalizations are related to particular observational facts. Of course "principles" and "policies" cannot be stated with the stringency and definiteness of "rules" in the narrow sense. What further interesting developments readily occur to us as springing from that conjectured analogy? First, moral reasonings that you and I learned in our happy positivist days to call pejoratively "circular" become less obviously so, and perhaps not circular in a vitiating sense at all. Given everybody's agreement that phenomenalism won't wash, an element of harmless, unavoidable "circularity" exists in common life and in science, at least in this sense: One is not simultaneously asking the same rock-bottom epistemic question about every proposition p belonging to a set P that may be involved in answering a legitimate epistemic question about a particular proposition q that is momentarily the object of our interest. While we would, even today, consider an argument—whether deductive or inductive—viciously circular in which one and the same statement S plays a critical rôle in a "proof" and S occurs in the intended conclusion, we do not say that all statements of a given *kind* are forbidden to appear somewhere in the body of evidence adduced for a probandum that is of that same *kind*. Thus, when we consider protocols that are provisional candidates for the corpus of empirical knowledge, whether in scientific or ordinary life, law courts, or whatever, we make it a practice to reject rather quickly those candidate protocols that are presented by "totally untrustworthy observers," such as pathological liars, psychotic individuals and the like. But everyone knows that we go about showing that somebody is untrustworthy or careless or crazy by showing, among other things, that he often reports protocols that we cannot allow to be "Neurath-eingliedert." That is, we do not say that it is viciously

circular for us to calibrate certain observers, as when a witness's testimony in a trial is impeached by successfully showing that he has lied or malobserved or misrecalled facts f_1 and f_2 , and therefore we hope the jury will not give much weight to his testimony on important fact f_3 , about which we have less direct evidence, conceivably no other direct eye witness evidence. The idea of *replication* in science can be dealt with in this way rather than by the usual positivist reference to "intersubjectivity," as if the latter were some kind of rock bottom condition for truth. (Here I agree with Skinner that it isn't. Robinson Crusoe, all alone on his island before the appearance of Friday, could have constructed and tested an adequate science and technology, without asking the modern empiricist's favorite question, "Will other people see the same thing I do under these circumstances?") Of course even if somebody is well calibrated, we may still decide against Neurath-inclusion, as happened in the famous case of Dayton C. Miller. He was a highly regarded and extremely careful experimentalist, a President of the American Physical Society, but his ether drift claims in the 1920s and 1930s were simply set aside as "occult effects." Within the last decade some careful and terribly complex computer analyses of the time cycles in his data corroborated one of the several conjectures that had been offered much earlier, namely, that his ether drift results were artifacts of temperature fluctuation in the apparatus. Now I was taught as an undergraduate that it was a *petitio principii* fallacy to say that a painting or a piece of music is probably of good quality because it is preferred by persons of developed taste, when the only grounds we have for saying that somebody had developed taste was that he liked such and such kind of painting and music (see, e.g., Alburey Castell, *A College Logic*, page 21). It is no doubt necessary to instruct the undergraduate mind that there is a grave danger of this circularity in such arguments; but once having done so, one should then go back and examine whether they are always and necessarily viciously circular. The problem is somewhat like the problem of identifying a disease entity from symptoms when we have to bootstrap the validity of a particular symptom on the basis of provisional classification of individuals, and then we realize that the way we classify the individuals provisionally (for use as a best-available criterion of symptom validity) is on the basis of all the symptoms, maybe even this one included in the batch! If one assumes—as I have throughout this memo—that some still to be explained combination of the coherence and correspondence theories of truth is going to be required *not only for ethics but for science and all other cognitive enterprises*, then there is something to Feyerabend's remark that "There's nothing wrong with arguing in a circle, provided it's a big enough circle." (C. I. Lewis said it first, in his great 1929 book, *Mind and the World-Order*.) As an undergraduate, I thought this remark fatuous; but I realize now that it was insightful and maybe even profound. In asking myself whether Jones is or isn't really an anthropologist, I don't find it objectionable to ascertain whether other anthropologists call him an anthropologist and accept him as one; but of course that means I have to have some basis for deciding whether *they* are anthropologists or not. There are very interesting questions here, such as defining smallest coherent subsets of persons identified by the fact that each

of them calls each of the others an anthropologist, even though there may be other people who call themselves anthropologists that this group does not, and so on. In ethics, if I know somebody is a scoundrel—a liar, thief, promise breaker, and generally callous, insensitive, hurtful individual—I will not include him in my set of “ethical intuition providers” when it comes to seeing what kind of an axiology I can articulate on neo-Prichardian “criteria.” I find it hard to think that even an avowed ethical relativist, if any really exist in this world now that Nietzsche is dead (and he wasn’t one, really), would think that it was a fatal criticism of a proposed axiology that it did not accord with the intuitive deliverances of John Dillinger or Adolf Hitler. That’s why people from way back, and writers like Prichard and Ross and Brandt, have made it a practice to refer to people’s “reflective ethical intuitions” about concrete cases. What is the structure of that kind of argument? You begin with an individual whose ethical sensitivities are not totally blunted, this judged by your own lights—just as the jury decides whether the witness is stupid or crazy or a careless observer. You ask whether this particular individual who has been, as a person, more or less “calibrated” by you ethically, is “an ethically sensitive and reflective person.” This does not mean he had to agree with you about *every one* of your ethical intuitions, or even about your major ethical postulates, any more than we assume that if Jones is a “more or less trustworthy witness” whose testimony should be given weight, that commits us to saying that Jones has never malobserved or misremembered or inadequately reported anything at any point in his testimony, let alone any point in his entire speaking life; nor that we have never malobserved anything that Jones observed correctly. We then ask whether Jones in this particular situation has given it thought, whether he is momentarily under the influence of some kind of passion, or has had too many bottles of strong beer in the course of the discussion, or whether we are tromping on his Ku Klux Klan or slob-lib or Marxist or Hegelian or Freudian toes, whether he seems to be listening to what we say to him and taking our discourse seriously and shows genuine rather than pseudo reflection when knitting his brows, and all the rest of it. All this kind of evidence is used, and, I want to argue, *validly* used, despite the possibility of abuse in committing a circularity of our own against Jones, in determining whether Jones’s “candidate ethical protocols”—if I may use that language here—are to be taken seriously. Finally we ask whether they are to be taken *very seriously indeed*, in the sense that we ourselves, on reflection, find that we strongly tend to share them, which would mean that an axiology which doesn’t do justice to them by subsumption, given the preference rules and so forth, has something strong going against it. Note that I do not say “has now been refuted,” let alone “must be abandoned,” for Lakatosian reasons applied to ethics. It may not have been refuted because there are some further considerations we haven’t thought of; it may not have been refuted because we may yet come around to abandoning our ethical intuitions *in concreto*; and finally, very importantly where ethics is concerned, it may have been refuted in the strict Popper₀ sense and yet we might not abandon it. We sometimes stick with scientific theories that are, strictly speaking, demolished *modus tollens*, and I do not see why we may never stick with similarly

“refuted” theories in ethics, when ‘refutation’ \approx discordance with an ethical intuition *in concreto* that we find ourselves having grave difficulty to abandon.

So much for the circularity of appealing to “ethically sensitive and thoughtful individuals (ourselves presumably included)” in testing an axiology against ethical intuitions *in concreto*. There is a related problem tied in with explicating how the protocols collectively control the network and yet the network is sometimes (often?) used to control the admission of a single candidate protocol. The related problem is to specify and justify the “psychological” (broad sense) properties and relations that we look for in a strong *prima facie* candidate protocol, in addition to the overall calibration of the individual who utters it as an intelligent, informed, rational, ethically sensitive and reflective person. I don’t know what the analogy here is to the scientific protocol case, although it seems there ought to be one. And if there isn’t one, an analysis of *why* there isn’t should shed further light on the similarities and differences between axiological and nomological network testing. I have said above that the proper way to examine the merits of a proposed axiological is in the light of the purpose of axiologicals which is to tell us how to act, by analogy with the purpose of nomologicals which is to predict and explain spatio-temporal events, i.e., to explain facts. For this reason the connection between the genuineness (or weight, or sincerity) of an ethical intuition *in concreto* to the strength of one’s dispositions to act in such and such a way, or to feel guilty after acting or failing to act in a certain way, might be considered more intimate than in the case of theoretical statements in nomology. Crudely put, I may ask, “Do I really mean it when say I have this ethical intuition *in concreto*?”, and I may then include, as part of the answer to such a question, the best available indications from introspection or from my own behavior as to whether the appropriate impulses and feelings are there or not. This might be analogized to asking a law court witness whether he saw the automobile license clearly, or the examining clinician how unmistakable the faint mitral valve murmur seemed to him when listening especially carefully through his stethoscope. But I don’t fully trust that analogy here.

I want to emphasize that I am *not* arguing that nobody can believe anything about an obligation if he fails to comport himself in accordance with it. I think that one of the strangest things that positivism ever generated was the claim by Carnap—I forget where it was but I’m pretty sure it was in print and not in conversations I had with him—that a person obviously could not genuinely prefer something or like something or want something that he considered wrong or bad. This is as bad as Socrates, except that Socrates was disadvantaged by not having available the Jewish-Christian tradition which has always emphasized the possibility, and in fact the ubiquity, of people zestfully doing things that they themselves know to be wrong. I am myself not a guilt ridden individual (although my wife says I am somewhat legalistic in small matters), but I find myself stupefied by such a bright man as Carnap making such a strange assertion as this.

So I am not, repeat not, arguing that a person doesn’t “sincerely believe in” an axiological unless he himself invariably conforms to it. I would, however, suggest the following weaker criterion to detect ethically incoherent, hypocritical, or poorly intro-

spected attitudes: Suppose we have an axiological A which person X—intelligent, educated, and normally sensitive, reflective, and articulate about ethical matters, and not momentarily driven by some intense passion or compelling desire—claims that he holds. And suppose we note that he usually departs behaviorally from what the application of the axiological would require. (I do not want to argue about extensional vagueness here, and I don't care for present purposes whether the axiological is of the rule type, principle type, or policy type, so long as it is sufficiently clear that it excludes some of the actions $A_1, A_2, A_3 \dots$ performed voluntarily by our moral agent.) More than that, we do not find ourselves dealing with a person who from time to time “slips a bit,” a person who, due to a serious personality flaw or weakness of the flesh, succumbs to intense temptation or acts impulsively when fatigued or overworked, and the like. Rather we have a situation where our moral agent X rather consistently behaves out of accordance with his purported axiological A. Further, we fail to observe in him the usual signs of guilt and shame (they may involve idiographic as well as nomothetic behavior indicators, so it may be necessary to know moral agent X rather well to use this criterion) that we have learned to discern he normally shows when he falls short of one of the principles that he “sincerely believes in.” Also, his subsequent instrumental behavior is lacking certain *undoing* features, such as what the traditional moral theologian would call “making reparation.” Nor does he tend to ask forgiveness or make apology. We discover, upon thorough exploration, that he is not apparently engaged in any sort of personal moral retraining or practice of the sort which would tend, according to his own degree of psychological sophistication and self knowledge, to increase his disposition to behave in accordance with axiological A. And, finally, he does not even show any higher-order disposition to be impelled toward acquiring such alterations of his own dispositions. That is, he lacks even what in Roman theology would be called a *velleity* towards self improvement. Summary: He neither follows Rule A, nor tends to follow it, nor tries to follow it, nor feels guilty about not following it, nor acts so as to make up for occasions of not following it, nor asks forgiveness or apologizes, nor takes steps to learn to follow it, nor experiences impulses to take steps to learn to follow it. What are we to make of such a situation? Granted that the weakness of the flesh can operate at the level of second order dispositions to reform one's first order dispositions; granted that even the more ethically sensitive, self observant, articulate and intelligent of us can at times engage in massive self deception, and can resist the interpretations and confrontations of our friends (or psychotherapists) on matters where we are strongly defended. Despite these alternative explanations, one plausible explanation of such a state of affairs is that the individual does not “sincerely, genuinely believe” the axiological A that he tells us he believes. (Here again, I don't want to say what Carnap once said to me in conversation about postulates, namely, that he who believes the postulates of course believes all of the theorems, merely on the grounds that the theorems are consequences of the postulates. I can see that one can speak that way, but it seems to me a misleading way to speak and one which has little to recommend it.) I do want to suggest that when a person lacks *any*

of the associated feelings and other dispositions that would normally go with holding an ethical principle A—especially one who is usually articulate and coherent and who, when failing to accord behaviorally with his own moral principles, normally does experience appropriate affect and then subsequent counteractive and reparative impulses—this pattern should be grounds for wondering whether his introspections about his intuitions are accurate or not. And, generalizing over the class of ethical intuitors, at least those within a given culture who are familiar with the standard classical gambits in this game, I think it reasonable to be suspicious of any axiological A, even if it is one defended “theoretically” by eminent persons and “officially” accepted by others who have read their work, if it turns out that the intimately connected network of allied dispositions to feel guilt or moral pride or to make reparation, *or to exhibit a velleity toward a second order disposition to alter one’s first order dispositions to behave conformably to it*, are uniformly or almost uniformly absent.

If I am right about this, or even if I am not quite right but my point does have some substance and non-zero probative value, then it seems to me that well over half of contemporary moralists are haywire. I have no statistics, but I think we can safely assume from the current literature as well as conversation among educated persons that utilitarians (including both act and rule utilitarians), and persons who hold some variant of G. E. Moore’s “ideal utilitarian” position, and the neo-Kantians probably make up 80 or 90 percent of the “ethically sensitive and articulate Western population” at the present time. I think these people are all ethically incoherent in the sense I have just spelled out, that their verbalized axiology badly mismatches their moral intuitions and associated dispositions *in concreto*. I leave aside Thomists and other Christians for obvious reasons. I also omit neo-Prichardians (like myself?) because we are an almost extinct species and, more to the point, because that is the one position, with whatever its other defects may be, that does not appear to me to be obviously refuted by what I’m going to say next. Let us take G. E. Moore. Although he was not a rigorist like Kant, and he specifically makes it a point to say that ethics cannot tell us precisely what to do (because nobody is able empirically to make the requisite computations concerning the whole, unforeseeable, indefinitely long chain of consequences that follow from a given act), he nevertheless ends up stating what we are obliged to do in probability, and also gives a very strong *prima facie* weight to follow a rule which, if we can show that its consistent or frequent violation would have bad consequences generally, we ought to follow unless we can make a terrific concrete empirical case that the contemplated “exception” will be different. By first intuiting what we mean by the adjective ‘good,’ and then having to intuit which things are good, we finally have to intuit that our moral obligation is, so far as we are capable of and so far as the general rule is formulable, to act so as to maximize the good. Now I want to urge that neither G. E. Moore nor any of those who profess to hold his view meet the above criteria of “ethical sincerity *in concreto*.” Take, for example, what one does with money. I believe that G. E. Moore was accustomed to having a couple of glasses of wine with dinner, although he did not abuse alcohol. At current

prices (I take it that inflation will come out in the wash in what I am about to argue) a glass of good wine—and knowing Moore and his background I think he wouldn't drink some cheap sherry but probably a high class port, port wine being his preferred example in his writings—would cost about 50¢, so that would mean a dollar per evening. That is \$30.00 a month, which is more than enough to feed a starving child in India or Guatemala. Now can it be seriously maintained that the utility of a couple of glasses of wine to G. E. Moore exceeds the utility to the child (and its parents) of life vs. death? We can soften the blow some by saying that the child won't die but that the various international charitable and relief organizations involved will merely have to spread their donations somewhat thinner; but that won't help much because rule utilitarianism will get us in the neck. And even on grounds of act utilitarianism \$30.00 less is \$30.00 less, and that means that *somebody* is going to be a *little bit more hungry*. It goes without saying that if this \$30.00 will support one life, the reduction of \$30.00 is going to make some individual child go to bed hungrier, or the parent who deprives himself or herself so that the child will not go to bed hungry. It doesn't really matter which, does it? Arguments of this sort would apply to anybody in an affluent society who fails to restrict himself to a diet of soybeans and wear one washable cheap suit. And obviously it is really atrocious for Meehl to buy a diamond ring or to own an air conditioner for his home. Similar arguments, with less dramatic extremes involved, can be made everywhere you look. It's not a matter of odd-ball, far-out "hypos"—the examples are all around us. Now the question is whether G. E. Moore and his disciples, or any act or rule utilitarian or neo-Kantian, feels guilty about this matter, or experiences a significant velleity to try to alter his attitudes and conduct so that he *will* feel guilty or have an appreciable statistical increase in occasions of foregoing a glass of wine so that an anonymous child in an impoverished country can have a glass of milk instead of going to bed hungry. I am, of course, aware that a saintly type, a type that most persons (and, I suspect, G. E. Moore and his set) would look upon as some kind of "moral fanatic," would say that this is one of the best examples we can find of the depth and pervasiveness of our sinful nature, since intelligent and otherwise ethically perceptive and sensitive persons can have a glass of wine without batting an eye realizing that they are thereby in effect "passively helping" to kill a starving child in India. I do not dismiss this argument lightly, being an ex-Lutheran and very much impressed with the folly and wickedness of mankind, myself included! But if we hold that it is at least *prima facie* evidence against the genuineness of a person's alleged ethical intuitions *in concreto* that he experiences a negligible impulse to behave in accordance with them (assuming the kind of mutual collective control between axiology and gut-level ethical intuitions *in concreto* that I have analogized to the interplay between nomologicals and protocols), then these familiar observations speak strongly against most of the ethical philosophies currently prevalent among educated Western minds. And I don't think that act utilitarianism is any better off or worse off than rule utilitarianism or neo-Kantianism or Moore's ethics in this respect. All these ethical theories are in bad trouble with the "sincere ethical intuitions *in concreto* of reflective persons," unless we

bite the bullet by insisting that the depravity of human nature leads to a dreadfully egocentric “out of sight, out of mind” psychology *vis-à-vis* the starving Hindu child and our glass of port. Again I say, I do not dismiss that Luther-Calvin-Augustine possibility. But if someone claims to hold it, I can only repeat, “Do you then experience a *velleity* to become otherwise? If not, are you sure, upon reflection, that you *sincerely hold* the Luther and Co. view of the matter?”

§ 25. Reflection on considerations raised by Dr. Edwin L. Crosby (President of the Jefferson Center, a local outfit on whose advisory board I serve) in his doctoral dissertation in Political Science suggests that our Prichardian ethical intuitions *in concreto*, which correspond to the “basic propositions” or “protocol sentences” of nomology, should perhaps be systematically extended to include (and with high weight?) the gut-level ethical intuitions of persons who have been “on both sides,” so to speak, of an interpersonal ethical relation. Crosby, in developing his ideas about what he calls “Extended Rational Discussion,”³ holds that I am in some important way disadvantaged as an ethical evaluator of, say, the welfare system, if I have never myself been poor, or low IQ, or ill, or out of work, or a widow with dependent children and no vocational skills; maybe I have never even known anybody as a friend or family member who has been in such a situation. One thinks here of the possibility of a kind of bootstrapping and calibrating, both of *persons* and *situations*, hoping to ascertain under what conditions and for what classes of ethical dilemmas it is possible for a fair-minded, empathic, and reflective moral agent to come up with ethical intuitions *in concreto* of the same sort that *he would be very likely* (we would not require certainty) *to arrive at if he were differently situated*. It’s a difficult problem and involves the same kind of questions that arise in criticizing Rawls’ theory of justice when he invokes the idea of rational choices people *would* make if imagining themselves to be in the “original position,” deciding behind the “veil of ignorance.” That is, apart from the criticisms that some (e.g., Brian Barry) have made as to whether the state of mind, or, putting it less psychologically, the state of knowledge and principles of reasoning, that must be attributed to the ethical decider in the original position behind the veil of ignorance can even be made coherent; assuming that these objections can be answered by Rawls, there remains the problem of whether I can really (empirically) judge how I *would* counterfactually feel about, say, a lexical ordering of liberty over and against average social utility or equality or whatever, when I have myself always been a relatively advantaged person in terms of IQ, verbal fluency, energy level, coming from a middle class home, and the like. It would be an interesting venture, for example, to study how well the reflective Prichardian ethical intuitions *in concreto* of persons of middle class origins and continued economic security tend to agree with those of persons born poor who have remained poor, as compared with persons born poor who have “made it” by a combination of brains, energy, and luck. And we would like to know whether the empathy of the “winners” for the “losers,” or the

³ [Cf. *ERDERVE* acronym, footnote 1, page 1.—LJY]

empathy of those born to well-off families for both the poor-born winners and the poor-born losers respectively, could be increased by indirect (vicarious, empathic) means, such as watching a television serial about a poor family. John H. Griffin dyed his skin and traveled around as a black American so that he could really experience what it felt like to be racially discriminated against. He concluded one simply could not empathize without experiencing it. But the idea, however one might try to implement it, is important, because one needs to know, in examining the candidate ethical protocols, something about the extent to which even the most intelligent, fair-minded, ethically sensitive person, upon reflection, is nevertheless psychologically *incapable* of intuitive convergence on what seems like “fairness,” “injustice,” “bad luck,” “disadvantage,” or “social cruelty” to the judgments of one who is and always has been in a situation economically, physically, intellectually, sexually, culturally, ethnically, or religiously totally unlike his own.

§ 26. When I read a collection like W. D. Hudson’s *The Is / Ought Question*, especially the papers dealing with what Max Black calls “Hume’s Guillotine” (the unbridgeability—I cheerfully admit—between ‘is’ statements and ‘ought’ statements), it is puzzling to me how little impact the developments in philosophy of science or even, for that matter, in general epistemology have influenced the reasonings of moral philosophers. Three examples of this stand out dramatically for me, but I suspect that careful search and reflection would discover others. The three presuppositions shared by almost all ethical philosophers who worry about Hume’s Guillotine, and—not identical with that but related intimately to it—the general stomach ache as to how one goes about defending or attacking ethical judgments, are the following:

- a. *Linear argument and resulting dreaded circularity*: We all seem to want our axiology to be capable of generating the “intuitively right moral deliverances” upon ethical dilemmas presented *in concreto*. But when we discuss *meta*-ethics, it seems that we feel methodologically sinful, because we can’t seem to escape Hume’s Guillotine; that is, we can’t seem to bridge the gap between ‘is’ and ‘ought,’ without committing a *petitio principii*. I admit the formal point, which (despite Searles’ infamous paper) seems to me clear—that *one cannot deductively reach a sentence containing ‘ought’ from sentences not containing this term*—but I am not overwhelmed by it. It reminds me of the well known business about arguing in a circle in artistic judgments, where one is not supposed to conclude, or even get support for, a thesis that Fisbee’s painting is a good painting by relying upon the consensus of competent art critics, and then explaining who belongs in the group of art critics on the basis of the fact that they, by and large and on the average and in the long run, “have good taste.” As I mentioned above, I see here in ethics, and maybe in esthetics (but I know nothing about that) a fairly close analogy—close enough to be illuminating philosophically—between the collection of protocol sentences on the one hand and the postulated explanatory nomological network of scientific causal theory on the other. Whether one is a justificationist and an inductivist or a Popperian or Lakatosian does not, I think, matter

much here. Because I believe it agreed on all sides, whether from general epistemological considerations about science or, more clearly shown, from study of the history of science and its common practices, that no simplistic linear derivation chain between a particular protocol (an utterance or engraving by an individual observer about a certain space time located observation or experience) and a scientific theory which is then, in turn, used to rule in or rule out another protocol is involved in any viciously circular way. What we rather have is an initial calibration, as I indicated somewhere above, of a bunch of scientific investigators on the basis of how well, in general, each individual performs and reports his experimental research. We rely partly on collateral evidence such as the testimony of persons who served him as colleagues or research assistants or whatever, but we also rely heavily upon how well his protocols have ended up, at least so far as we currently can tell (and we think maybe indefinitely), as capable of being fitted into the nomological network. Thus, Dayton C. Miller's ether drift reports, while they were put on the shelf until further notice because of the scientific community's acceptance of Einstein's theory on what seem very strongly corroborated grounds, did present a special embarrassment to physicists, because Dayton C. Miller had previously established a strong reputation as an extremely careful and accurate experimenter whose protocols were trustworthy. If I remember correctly, his report of finding a positive ether drift, *contra* Michelson and Morley, was actually delivered as his presidential address to the American Physical Society! While nobody, including Imre Lakatos, has given a completely satisfactory account of these matters, what I take to be generally admitted today, by both philosophers and historians of science, is that the collection of protocols in some collective sense constrains the explanatory theory (nomological network), but the "currently accepted on good grounds thus far" nomological network is often used to exclude individual candidate protocols from the corpus. In this respect Otto Neurath was right.

One might think in terms of "strong candidate protocols" as observational statements emitted by calibrated observers, the ideal case being that somebody has a degree in physics and has managed to get his paper accepted by a journal editor, even though we don't recognize the name and have to rely upon whoever trained him as an experimental physicist to be able to set up the apparatus, to be able to perceive the pointer readings and record them, and not to be a liar. That is, the candidate protocols have a kind of *prima facie* admissibility into the corpus, although that doesn't commit us to accepting each of them upon reflection, but merely puts them into the mass of "almost raw data" by which the nomological network is evaluated. So that protocols reported by paranoid schizophrenics (most of whom would be judged such on the basis of collateral evidence, although the content of their protocols about other matters than the one under consideration is usually part of that evidence) are not even serious candidates; and we do not worry about the anomalies in the ongoing research program, or even the specific substantive theory at this stage of the program, that would be created were we to admit these protocols into the

corpus. When a nomological network is currently received and has been well corroborated despite the presence of certain anomalies (I shall accept for purposes of discussion here the Lakatosian view of these matters, which, despite its defects, and his lamentable death before he had time to fix it up, I see as the most progressive research program going in philosophy of science), some of the candidate protocols are nevertheless not includable because they are theoretically anomalous, and these (because of their auspices, e.g., whose laboratory they came from, or the fact that they were replicated by several observers in different labs) we “put on the shelf,” hoping that sooner or later they will be explained—as was the Dayton C. Miller ether drift business, although it was “on the shelf” about 40 years before it was satisfactorily explained! In thinking about science in this way, we do not worry about a kind of simple minded “circularity” that somebody might allege if he didn’t pay attention to how complicated the network is and to the fact that we deal with a collection and treat the protocols as a mass, so to speak, or collectively rather than distributively. That is, somebody might say, “This is all terribly circular, because you decide whether you put in an observation or not on the basis of a theory, and then you turn around and tell me that your theory is a good theory because it fits the observation—obviously a grossly circular proceeding, shame on you.” Nobody says this. Why not? Because the situation is more complicated than a single protocol “leading to” a single theoretical sentence which then leads back to that same single protocol. And yet when we talk about ethics there is a tendency to worry about that kind of simple, linear inference chain circularity. My view is that we will have “calibrated ethical intuitors,” persons of intelligence, education, having a reasonable batting average for thinking clearly, not known to be political or moral fanatics, not considered nutty on other clinical grounds, familiar with the moves in the ethical game or at least open to be familiarized to them, and finally, whose ethical intuitions *upon reflection* (I have to keep using Prichard’s favorite phrase, despite his tendency to use it high-handedly at times) agree with our own and with other people belonging to the same initial *prima facie* ethical judges. Of course there are extremely difficult problems here about cutting scores, open concepts, differential weights, etc., and I take it that a developed ethical econometrics will in its pragmatic side include methods *such as nobody today can even conceive* of calibrating the ethical sensitivity, articulateness, and rationality of persons. In other words, I am a “conceptual optimist” about ethics. I view us today, as I said above, as operating in an extremely primitive state cognitively in this area. If somebody objects that I must choose between evaluating gut-level ethical intuitions *in concreto* on the basis of my axiomatized and numerified axiological network or, alternatively, evaluating the network by the intuitions, but that to do both by means of both is “circular,” I give him my analogy to the scientific case. That is to say, imagine an axiology that fails to jibe with a sizable subset of strong “ethical protocols” (i.e., the ethical intuitions that persons in the intelligent, articulate, ethically perceptive and sensitive class find

- themselves unable to abandon even upon reflection, upon putting themselves in one or another artificial way into the situation of the disadvantaged person who is challenging such and such a legal or moral rule, and the like). Then that axiology, to that extent, has anomalies which must be taken seriously. I shall say something about what we do with such anomalies in Subsection (c) *infra*; the point here is that there is no vicious circularity involved if you say it right, that a satisfactory ethics must fit the deliverances of the “moral sense” *in concreto*, and yet we sometimes correct or modify or even try to change the attitudes involved in a concrete moral response on the basis of an axiology which is, on the whole, doing well enough for us to continue using it.
- b. *Justificationism*: When we think about ethics, and very much when we think about meta-ethics (I guess today you could say most moral philosophers don’t even do that—they think about *meta*-meta-ethics), we seem to take it for granted that our task occurs within a general epistemological framework of justificationism, as if Sir Karl Popper had never existed. For that matter, one does not need to mention Popper in this connection. One could equally well point out the epistemological despair (within a justificationist framework) stated, if I recall rightly, quite explicitly by Lord Russell in his later writings, where he points out that one cannot legitimate inductive enterprise (of course I don’t merely mean the straight rule here) without certain metaphysical presuppositions that either stand totally unjustified or are defended by some kind of circularity that seeks to elude Hume’s problem. Thus, for instance, I am reasonably satisfied that no “logic of inductive inference,” however souped up and prevented from generating contradictions of the sort that Popper has pointed out, can quantify probability or statisticize confirmation or even generate a rough ordering relation among competing theories or among theories in different domains, without somewhere implicitly assuming such things as Keynes’ Postulate of Limited Independent Variety and Postulate of Permanent Kinds. If all “generator properties” are independent of the others and (only denumerably) infinite in number, or if the basic kinds of which the universe is composed for some crazy reason (e.g., intrusion of Satan or Mark Twain’s God as the Great Jokester) changed the supreme dispositions (C. D. Broad) of all natural kinds, then our efforts to understand the world, or to forecast it, would inevitably fail. Of course, like you, I don’t think that Sir Karl delivers affirmatively in the sense of showing why anybody should “reasonably continue to rely” upon theories merely on the grounds that they have as yet escaped refutation, apart from the incoherencies in his position concerning the virtue of *modus tollens* vs. the virtue of “scientific tenacity.” I think Lakatos has shown how to fix that one up, but the last time I talked to Imre before his death he said that he had become convinced, much against his will, that it would never be possible to show why it was the part of reasonableness to rely upon the hard core of a progressive research program. But I didn’t mean to wander off into all of that—I merely wanted to indicate that one does not have to be an orthodox Popperian (if there is such a thing) in order to say that any

simple minded kind of inductive justificationism won't wash, and that such a bright man and careful thinker as the late Rudolf Carnap wasted the last 20 years of his scholarly life on a degenerating research program. Point: We do not say, or at least I do not know anybody who says, "Show me how you *ascend to* the theory of the gene, without conjectures or postulations but by simple abstraction of the phenomena of mating and the resultant offspring phenotypic traits?" Everybody today, whether he has come strongly under Sir Karl Popper's influence or not, recognizes that science makes bold conjectures and then tests them. If somebody says to me, referring to my 1962 theory of schizophrenia, "But how do you arrive at these ideas?" my standard answer is, "I make them up out of my head—where do you get your ideas from?" But when we come to discuss ethical postulates, it seems that we take for granted a kind of old fashioned simple minded justificationism. We keep thinking it ought to be possible somehow to pile up the edifice of axiological statements upon a "foundational" set of empirical observations. If that seems an exaggerated criticism, I invite you to re-read some of the passages in the Hudson collection. Scholars don't seem satisfied to say that we are entitled to postulate primary or fundamental axiologicals, that we don't have to lift ourselves up into the axiology from some kind of inductivist justificationist data base. In ethics, *Hume's Guillotine hangs over us in justificationist form*. It may be that I am really addressing myself here to the same issue as I did in subsection (a) immediately preceding. In some deep sense, the "linear circularity stomach ache" and the "justificationist stomach ache" seem to be, at rock-bottom, the same problem. But I am not at all sure of this, and therefore I want to list it separately.

Of course, somebody may say that even a devout Popperian considers he has to test his freely made, invented, created, conjectured postulates in the nomological domain by going down to the facts. Whereas we have *not* shown what is the "analogous" operation in axiology. But that's a separate objection. The first form of the objection is the simplistic complaint, "You have not justified your primary axiological postulates"; and what I am pointing out here is that I don't see why I have to be a strict upward seepage justificationist about ethics when I have ceased to be a strict upward seepage justificationist about physics, genetics, or psychology. Not that I thereby repudiate an obligation to answer the next question. But I have tried to indicate roughly how I would go about that, namely, by checking the consequences of an enriched, articulated, metricized, axiology against the strong reflective deliverances of the moral sense, of our Prichardian intuitions *in concreto*.

As I discussed in an unpublished paper you read several years ago, even if we say that one can postulate any ethical primary axiologicals he pleases and then "see how they work," *one* obvious meaning of "working satisfactorily" is being coherent, and that includes not only internal consistency (from the formal and semantic stand-points) but also generating consistent moral advice in fact situations. That is, even before we inquire whether our reflective Prichardian judgments *in concreto* are compatible with the consequence of applying a certain theorem that flows within the

axiology, we can investigate whether the postulated axiology, *when conjoined with well corroborated empirical statements about mind and society*, yields incompatible particular moral judgments. Since the “pragmatic function” (see Rescher!!) of ethics is to prescribe conduct in morally significant situations, just as the pragmatic function of postulated theoretical nomologicals is to explain and predict spatio-temporal events, when there is a failure to prescribe conduct—which is obviously the case when act *a* and contradictory act $\sim a$ are both prescribed for one and the same agent in a particular moral dilemma—then, to this extent, the system has “failed to perform its function.” That is, a prescription of incompatible conduct in a moral situation is a failure of an axiology, analogous to the misprediction of a fact of observation being a failure of a nomology. I repeat, this kind of incoherence is different from an internal incoherence based upon an inconsistency in the axiologicals themselves. It is an inconsistency in a somewhat more interesting and complicated sense, that it generates incompatible moral prescriptions when conjoined with empirical non-axiological statements. Of course the only way it can do that is by virtue of the fact that the axiological statements include descriptive predicates. But what makes an axiological statement an axiological statement is not that it is free of “naturalistic” descriptors, but that it contains one or more essentially axiological predicates such as the word “ought” or, if one is a follower of G. E. Moore, a word like “good.”

[Here again, there is an interesting analogy to theoretical statements in causal science, namely, that in order to generate consequences which permit us to refute nomological conjectures, while some of the nomologicals may contain only theoretical terms, some of them must contain a mixture of theoretical and observational terms, such that by putting these together we generate sentences from which the theoretical terms have dropped out leaving only statements containing logical constants and operators and mathematics plus observational terms. But I am getting more into the details of the analogies and disanalogies than I want to for present purposes.]

The basic point is simply that one who has been “softened up,” whether by the collapse of Carnap’s *Aufbau*, or Bertrand Russell’s late stage metaphysical pessimism, and surely one who has been influenced to any appreciable extent by Popper’s critique of justificationism and inductivism, ought to carry over some of those lessons into his thinking about matters ethical. He ought not to feel obliged to provide an answer to the justificationist question, “From what empirical facts, from what observational statements about natural predicates, do you ascend by inductive syllogistic inference to your axiological postulates?” If anybody puts the question like that, or relies upon an argument which implies that he is entitled to put the question like that, it seems to me we ought to say that we don’t admit a justificationist epistemology in ethics, any more than we admit one in science.

- c. “*Instant-rationality*” *perfectionism*: I remember how amusing Imre Lakatos was the first time he visited the Center, when he was talking about the (perhaps mythical?)

“Popper₀” as being, or seeming in some passages to be, a believer in “instant rationality.” I recall how Imre, in talking about a naive falsificationist position, would pass his finger across his own throat as if he were cutting his throat when mentioning a theory being immediately and definitively slain, liquidated, *modus tollens*. It seems to me that many discussions about ethics proceed in this fashion. That is, the fact that a proposed ethic generates an unpalatable moral deliverance about a certain situation (particularly a far out hypothetical, as the lawyers call it) and that our discomfiture when we reflect upon that particular moral deliverance refuses to go away, constitutes a definitive liquidation *modus tollens* of the proposed axiology, so that one ought to drop it and not consider it further unless he can come up with an adequate solution to the difficulty. It seems strange that we should impose such strict falsificationist meta-criteria in ethics, when we have been forced to abandon them for chemistry or astronomy—not to say psychology or economics! There must be hundreds or thousands of examples in the literature of moral philosophy in which somebody’s ethical system (or, more commonly of course, what the critic conceives to be a consequence of somebody’s meta-ethical principle or principles) is attacked by showing that it would entitle me to crucify an assassin or require me to let a paranoid schizophrenic blow up Manhattan Island, or to treat the Martians as if they were animals, or to reinstitute chattel slavery or the like. Now this point should perhaps be split: (1) the requirement that an adequate axiology must cover all imaginable cases, including cases that have never arisen and probably never will; (2) the point that it generates contradictions in some actual cases, which we cannot see immediately how to resolve by means of some kind of preference rule (although I assume many or most of these would be resolved by a more axiomatized and numerified quantitative axiometrics); and, finally, (3) the mistake (which Lakatos pointed out Popper had made in moving directly from a *modus tollens* refutation to the prescription for scientific abandonment) that if refuted, then abandoned. The last is a mistake in that, even if one admits that such and such a counter instance is a *modus tollens* refutation from the standpoint of Popper₀ “instant rationality,” that doesn’t immediately tell us that we have to abandon or junk the axiology, any more than it tells us we have to junk quantum mechanics just because we can’t figure out the Schrödinger Cat Paradox, or can’t satisfactorily explain the two slit experiment. One has the impression of a universal assumption on the part of meta-ethical theorists that if you can really make a good case that so-and-so’s proposed ethic will not wash when confronted with your (or preferably his own) ethical intuitions *in concreto*, then he is utterly destroyed. But all we have really done, as in the case of *modus tollens* scientific experiments, is show that something needs to be fixed up, though at the moment we can’t quite see just how. And the same is true of “gaps,” that is, ethical dilemmas which the axiology leaves open.

§ 27. There is some deep stuff here, “deep” both in the senses of murky-obscure and fundamental-pervasive. Once we reject, as I think we must, a *wholly* Ramseyfying

upward-seepage theory of meaning for the empirical sciences, something will, one hopes, come to replace it. Might we not use that something also in ethics? I am vague about it, and this memo is already far too long, but I want at least to record the core idea (however crude and fuzzy) lest we forget to work on it further. Consider a theory in physics, say our old favorite paradigm case, kinetic theory. The positivists' metatheoretical account of how the meaning of "velocity of molecule" gets specified (introduced? defined?? explained???) began with a kind of direct verificationism (American variant: Bridgman's operationism) that was soon seen to be untenable. Then came Carnap's 1936-37 attempt via reduction pairs—where "reduction" already ceased to mean "explicit definition without loss of meaning," and where the unsatisfactory (but truer) idea of *partial meaning specification* was invoked. Later we had a still more elaborate (and more accurate) approach, where an uneasy partnership was formed between the metamathematician's formalist program of "implicit definition" and the modified operationist view applied to a *proper sub-set* of concepts via "coordinating definitions," "meaning-postulates," or "bridge-laws." This concoction could, it seemed, be blended by means of the Ramsey Sentence, whereby the implicit definition notion from the formalist program of metamathematics is somehow "empiricized" by the special rôle of those terms that remain after Ramseyfication, because they belong to the "observational's" vocabulary. So here the upward seepage theory of empirical meaning is given a semi-rigorous metatheoretical expression. You may recall that on one of our visits to Carnap I tried (vainly) to persuade him that a purely upward seepage theory of meaning would not quite do—for reasons allied to Campbell's famous discussion of the "dictionary" and Mary Hesse's analysis of models. I said then, and I say now, that one understands the "notation $(\frac{1}{2}mv^2)$ " (for the single hydrogen molecule) in kinetic theory only partly by upward seepage. The Braithwaite analysis of a theory taken first as a pure ("uninterpreted") calculus, where the only non-formal components of meaning subsequently arise because a partial subset of terms are coordinated to observation predicates, is part—perhaps most—of the story; but it's not the whole story. The embedding text of a physicist's derivation of the gas laws from kinetic theory gives part of the meaning to v , by a carry-over from our acquaintance with the velocities of macro-objects like cars and billiard balls. A stronger claim (as I showed during early Center days for Estes' 1950 derivation of the acquisition function from neo-Guthrieian learning theory postulates) is that even the formalism won't usually go through without the embedding text. In Estes' famous derivation, only one of over a dozen transitions in the formalism can be made from mathematics alone—and even that one, integrating his penultimate differential equation, requires his text to interpret (and thereby evaluate) a constant of integration. I believe this is true of almost all "derivations" I have ever read in physics, chemistry, or genetics. So it isn't merely that the embedding text narrows the "admissible interpretations" of a calculus, or that it "fills out the meaning" somehow (beyond that data-base tie that "anchors the theory in observations," as the old Hullian language said it). Even this anchoring requires valid derivations from the theory-language statements to the observational consequences,

and these derivation chains require the non-formalistic text, contrary to the Braithwaite picture. (For examples other than my 1955 hatchet-job using Estes, see any physics book, e.g., derivation of the ultraviolet catastrophe from Jeans-Rayleigh model, or Millikan's long and many-principle deduction of why e could be estimated by the oil-drop method.) Summarizing dogmatically, I say that the meaning of theoretical terms like "molecule" and [molecular] "velocity" is specified for us in a complex way, by combining

- a. Implicit, rôle-in-calculus definition
- b. Upward seepage from the observational theorems
[Where (a) + (b) = Ramseyfication]
- c. Analogy-model-carry-over in the embedding interpretative text
that is *not* coordinated directly.

Thus terms like "particle," "distance," and "angle" partake, as applied to theoretical micro-entities, *in varying amounts* of the meanings they originally received from our childhood learning and adult thing-predicate usage of language about macro-objects. And, as Mary Hesse has shown (Hempel agreed at one of the big Center conferences), this partial carry-over of meanings cannot be relegated to mere "context of discovery," as Howard Kendler tried to do in his paper "What is learned?—A theoretical blind alley," which Rozeboom clobbered in a Center seminar.

Well, what has all this to do with ethics? I don't know exactly. But I'm pretty sure it has something to do with the question, "What do ethical predicates (or obligation-vectors) *refer to*?" We do not want to psychologize obligation, as is done by those who try to introduce ("define") it in terms of such entities as pro-attitudes, commands, exhortations, or avowals (whatever the heck they are—I've never quite grasped that). But if we reject that approach, how then does it come that we nevertheless make a connection—and candidly say that we feel we *must* make a connection—between the souped-up normative calculus and our reflective (and collective) Prichardian moral intuitions *in concreto*? We do not straightforwardly *mean*, by a statement imputing an obligation vector $O(x, y, a) = 12$ to moral agent x that he perform act a for person y , that we "feel a pro-attitude of 12 introspectible obs." (Efforts by moral philosophers to *reduce* obligation statements to a mental state are, on my analysis, misguided in the same way that verificationist, phenomenalist, and operationist "translations" of scientific nomologicals are misguided. He who tries to explain an axiological by a direct and simple tie with somebody's pro-attitudes, or disapprovals, or guilt-feelings, or over-riding impulses, or reflective intuitions, is making the *Aufbau* mistake in ethics.) But if such a straightforward reduction is not in order, then what *is* the connection between Prichardian moral intuitions *in concreto* and "theoretical axiologicals"? My hunch, to be explored later, is that some component of moral meaning is thereby contributed, that the ethical theory-word 'ought' which appears in the axiomatized and numerified utopian axiology is analogous to 'velocity' in the kinetic theory. The 'ought' in an axiological network gets part of its meaning by its formal rôle; it is also somewhat Ramseyfied (because the

network “applies to” situations, it classifies actions as licit, obligatory, forbidden). But a pure thinker who had never experienced even a smidgeon of the distinctively moral sentiment, who had never felt obligated, or guilty, or blamed, or judged—such a person would not understand quite what the word ‘ought’ is about. Analogy: Imagine a pure intellect which had *logic* and *number concepts* and which had *experienced change* (but only in the form of altering hues in a *Ganzfeld*, never as motion, displacement of a seen object in space). Also this knower senses the temperature modality and perceives that there are different temperatures at different times. Would such a person “fully understand” the kinetic theory? Would the meaning of v receive the meaning it has for us, on the basis of Braithwaite upward seepage in the “partially interpreted calculus”? I think not. I think I had a point against Carnap, and I think an analogous point holds for ethics. An essential component of the meaning of ‘ought’ is provided by the distinctively moral conations and affections we experience *in concreto*. The mistake is to *reduce* the ‘ought’ to these psychological events, states, and dispositions. I think the theoretical ‘ought’ is deeply connected to the moral dispositions, but I can’t say how. It’s rather like the bridge-law between kinetic theory’s micro-quantitative variables (momenta mv of hydrogen molecules) and the macro-quantity (a “pressure” on the piston). Feyerabend always argued that so called bridge-laws (worse, meaning *postulates*; still worse, coordinating *definitions*) should be theorems, and you recall that I took that position in the mid-1950s, before I had met him and got ammunition from him. One reason we find kinetic theory “intellectually satisfying,” the reason we feel that it really *explains* the gas laws (and the thermometer’s readings!) is that it is not a phony pure-calculus derivation of the sort Campbell concocts (to poke fun at it), but rather that the “bridge” between notation Σmv and macro-pressure on the piston is a *non*-arbitrary, *non*-definitional “bridge,” a bridge that seems built inexorably because we already have an interpretation of the theoretical notation ‘ m ’ and ‘ v ’ that is, to use Freudian terminology, anacritic upon our ordinary pre-scientific mass and velocity language for macro-objects. I realize I am not putting this very clearly, because I’m not very clear about it. I’m only pointing vaguely at something that I’m pretty sure is there to be explicated. “Ethics as pure postulate” is (partly, not wholly) meta-explained in terms of *commitment*. The rules are about how to act, so we should expect that one essential component of the meaning will be affective-conative rather than “purely indicative” (if there is such a thing for embodied knowers). Suppose a “pure knower” was presented with a system of moral rules, such that all the predicates except one—the ‘ought’ relation—were reducible to psychologese. Thus, terms like ‘promise,’ ‘cruel,’ ‘disappoint,’ ‘reward,’ ‘equal pleasure’ are all psychologized; and we then go on to reduce what at first might seem distinctively axiological terms (e.g., ‘justice,’ ‘due process,’ ‘exploitation,’ ‘benevolence’) to these previously psychologized concepts. I do not assert that this is possible, but if not, it strengthens the argument I wish to make. Thus we assume all of our quasi-ethical vocabulary to be reducible to psychologese, all apparent prescriptors reducible to descriptors, except for ‘ought’ (or, better, except for some such word or phrase that provides part of the embedding text, and hence

a partial interpretation, for the formalism of obligation-vectors). Finally, assume that the generic concept of a *rule* can be adequately explained to such a knower. I rather doubt this, but let's assume it. At least I think the concept of a *rule* can be freed of a "strong obligatory" component (e.g., it is not reprehensible—although it is pointless—to cheat at Solitaire). I sometimes have the fear that 'rule' is a primitive term, but let that pass. We are assuming that 'a rule' is explainable, and the embedding text contains the distinctively axiological predicate 'ought.' Now what is our pure knower's understanding? He is a non-actor, he never does or suffers anything, he is never rewarded or punished, praised or blamed. Like my murderous cat Schrödinger (box score this summer: 16 birds + 14 moles), he has no Super-Ego. But he does grasp the distinction between a *rule* and an *empirical (descriptive) generalization*. Core of it is, for him (mebbe for me too?), that (true) counter-instances refute a generalization, but not a rule. The result of a counter-nomological event is that we classify the generalization as false. The result of a counter-axiological event is that we classify the event as a *delict*, and the agent as a "rule-breaker." (We do not, it seems to me, necessarily classify him as bad, wicked, weak, blameworthy, or deserving of penalties; although we typically add these things.) So our pure knower, I assume, is now capable of "applying the axiology," in the sense that he can sort actions into licit and illicit, into rule-conforming actions and delicts. (He can presumably further sort the rule-conforming acts into *merely* conforming and *rule-obedient*, but that important refinement is, I think, not crucial here.) Is there anything about 'ought' that our pure cognizer doesn't understand that you and I do?

The quick-and-easy positivist response to this kind of question is, of course, the same for 'ought' as it was for phenomenal hue words in liquidating the "empirical meaningfulness" of the inverted spectrum problem. I can almost hear you in the old days of the Center, when Wilfred Sellars distinguished between "Feigl MWF" and "Feigl TThS,"⁴ explaining to us what the identity thesis meant to you. You used to say (and I used to agree with you, until Wilfrid twisted my arm) that the experience of a yellow color rawfeel does not, strictly speaking, "find a place in the scientific causal network," and therefore a blind physicist who understood all about psychophysiology and physics and the retina and so on would, again speaking strictly, *know* everything that a seeing physicist would know. You will remember that it always bothered me to relegate the rawfeel quality of yellow to something called "mere pictorial imagery," because when we were talking about visual perception it didn't seem to me that there was anything illegitimate or somehow cognitively inferior about the subject matter being partly "pictorial" or "imaged." And I believe I convinced you in "The compleat autocerebroscopist" that one could make rather strong arguments—not merely by reiterating "I surely know what the yellow color looks like subjectively, whatever else I don't know..."—that he who has experienced a certain rawfeel event (whether the identity theory is true or not) has a definite cognitive edge over somebody who knows the language in all other respects but

⁴ [MWF is Monday-Wednesday-Friday, and TThS is Tuesday-Thursday-Saturday.—LJY]

is blind. At least the last time we talked about this, you said you found my two thought experiments on the subject persuasive. Of course I wouldn't admit that only knowledge useful for predicting future events is genuine knowledge, but the point of my two examples was to show that the unique knowledge a person has had when he has experienced a rawfeel of a certain quality would even meet that rather strong (and arbitrary) criterion of "genuine cognitive knowledge," so I am assuming you agree with me. This isn't directly relevant to the problem of the partial interpretation of the ethical term 'ought,' but I bring it up here because the positivist temptation in answering the question at the end of the previous paragraph is to say, "Our Martian *sans* super-ego, our pure cognizer who is never an agent or a patient, would lack no knowledge that Feigl and Meehl have—he would simply not be undergoing certain subjective motivational-emotional correlates of ethical knowledge." That is, I see an analogy (which I trust I use for expository rather than probative purposes—to "soften up" the residual positivist in Feigl and Meehl together!) between a hard-line positivist rejection of a seeing person's color hue experiences as being "mere pictorial appeal" and a hard-line positivist declassification of a Prichardian ethical intuition *in concreto* as having "no cognitive content, but merely psychologicistic emotional-motivational overtones." What I want to say (I repeat crudely and fuzzily and as a first approximation to a point that I believe we could fix up with some more work) is that, to the extent that an ethical or legal rule is a member of a subset of rules that contain some kind of prescriptive-judgmental-obligatory feature that is not present, or at least is present in a very attenuated and near zero amount, in the rules of Solitaire or the rules of etiquette or grammar (I must say I have never been clear about the rules of logic in this regard!), part of understanding how such language is employed by moral agents should, in the sheerly cognitive sense of fully understanding a language network, include the distinctively "obligatory" or "shouldness" component. And I want to argue that our pure indicative action classifier from Mars, where people have no power to benefit or hurt one another and nobody ever relies on anybody's promises because nobody ever does anything but cognize the flux of sensory events, would not fully know how the word 'ought' is used by us Earthlings, in somewhat the same way—I of course do not mean in exactly the same way—as the non-seeing utopian psychophysicologist would not know *fully* how the phrase "red phenomenal hue" was used by a seeing psychophysicologist. And I want to suggest, however clumsily, that the collection of Prichardian intuitions *in concreto* against which the overall satisfactoriness of a proposed axiometric system would be judged, plays two distinguishable rôles.

The first rôle of these Prichardian intuitions is easy. They are ethical judgments about concrete moral situations (the Jesuits' "cases of conscience") and can be expressed as propositions (that there may be a certain extensional and even intentional vagueness here we have learned to live with in physics and psychology, so why not in ethics?), and those propositions could, I suppose, be related to the overall axiological network by an amoral Martian in the same way that the blind utopian psychophysicologist can relate people's color perceiving protocols to the rich nomological network of physics, retinal receptor

chemistry, brain function, psycholinguistics, etc. The amoral pure cognizer Martian may have to have some kind of metalinguistic explanation which makes it clear (and I can't see why this would be terribly hard to make reasonably clear) how axiologicals differ in the Earthlings' use from nomologicals; namely, a nomological is supposed to fit all of the (accepted!) observational protocols, and is considered refutable *modus tollens* by an accepted counter-protocol; whereas, *per contra*, an axiological is not in trouble *modus tollens* when it doesn't fit the conduct of an agent, but rather the agent's conduct is classified in a distinctively axiological category known as a delict. It seems obvious that this could be explained to the Martian without telling him anything about how people feel when they do that which they ought not to do and so forth. Whether we have to explain the concept of 'rule' more generically I don't know, but I'll take it either way. I have assumed above that the notion of a rule as an expression which *classifies instances* rather than as a statement which *must be in accord with them* (like a purported law of chemistry) is something that the Martian understands, despite the fact that his cognition is restricted to purely indicative discourse; you don't have to have a moral sense or Super Ego or be an agent or patient to understand that a certain derivation in pure mathematics fails to accord with a rule, and I should suppose a fantastically rigged pure cognizing Martian would have the same amoral capabilities. Thus the first function of the Prichardian intuitions *in concreto* is analogous to the theory testing function of basic statements for Popper, in that the Martian can understand what an Earthling moral philosopher is up to when the latter is trying to fiddle with the generalized axiological network so as to put it into a better fit (I repeat, he does not require a perfect fit except as a regulative ideal) to the collection of Prichardian ethical intuitions *in concreto* that have been, until further notice, accepted as strong candidates for reception into the corpus.

Before I pass on to the second function, let me point out that this is, from the purely cognitive-indicative standpoint, a rather happy state of affairs, despite logical positivism. By far the greater part of the Earthlings' ethical discourse would be comprehensible to the amoral, pure contemplative, indicative-tokening Martian. We can easily imagine the amoral Martian and a pious Jesuit moral theologian having a perfectly intelligible conversation about one of the famous "cases of conscience" found in Jesuit manuals of moral theology. I don't think this should strike us as any odder than the fact that a rational being from Paradise could have conversations with Keynes or Pigou about welfare economics, despite the Paradise dweller's failure to have experienced scarcity of goods or competition between persons for goods—whereas all economic texts I have ever read begin by saying that economics even *exists* as a discipline only because of the empirical fact of scarcity. Air, for instance, was formerly used by economists as an example of something that is not a "scarce good"; but now that we worry about pollution, air has come under the purview of economics. I cannot resist the impulse to point out that, here again, it seems strange that such a bright person as Freddy Ayer could have been so high-handed about the non-cognitive meaning of ethical statements in his 1936 book, given the possibility of a rich, complex and highly rational conversation of this sort.

And, before going on to the second upward seepage function, I must also add that it would not be terribly distressing to me if this is what we were stuck with. That is, if all the rest were relegated into “impure pragmatics” (whatever such a relegation means today) one would still have a discipline of ethical philosophy, would he not?

It is not clear to me whether all, or nearly all, of what is involved in the application of a weakly prescriptive rule (i.e., a rule that has a zero or negligible legal or moral loading, such as a rule of Solitaire or crossword puzzles or grammar) can be expressed equally adequately in either object language or metalanguage discourse. My hunch is that some kinds of things we want to express by enunciating rules can be done either way without loss of intended content, but maybe others can be done only metalinguistically. It strikes me that this question may be similar to the logic of probabilities, where those probabilities that are inherently epistemic in nature, such as the probability of a hypothesis upon evidence in a court of law, are inherently metalinguistic because they purport to relate one proposition, the probandum, to other propositions, the evidence; whereas those probabilities that occur as probability-numbers in the object language (e.g., in genetics, psychometrics, theoretical physics) *can* be kept in the object language of the particular science, but they also lead under certain circumstances to epistemic statements for which the same probability numbers may be appropriate. In the case of obligations, the net obligation vector in an action space would be, in our axiomatized and numerified utopian numerical axiology, object linguistic in the same way that statements in econometrics are (usually) object linguistic. I don't want to get into a general discussion of the nature of a rule, although at the moment I have the impression that to say what a rule is may be a little easier than is usually thought. If the essential feature of a rule that distinguishes it (when minimally prescriptive) from an empirical generalization is that we do not countenance the notion of its being “refuted” or “destroyed” or “invalidated” *modus tollens* by the occurrence of a counter-instance that is accepted into the corpus of empirical statements, then it doesn't seem to me that it ought to be so terribly difficult to explain to a purely cognitive non-agent non-patient indicative-tokening Martian how Earthlings use rules. My usual test in these matters is to ask whether I could get a human being who, like myself, does have an intuitive preanalytic understanding of a term, to come out with that term after I have presented him with a set of statements about the entity the term designates, each time replacing the term I want to elicit by a Greek letter. Thus, for example, it is interesting in the Philosophical Psychology class when I provide a progressive “contextual definition” of the clinical (or learning theory) term ‘anxiety’ by saying things like: “There is an organismic state α whose decrement constitutes a positive reinforcement.” “The state α is elicitable by conditioned stimuli.” “The state α has as unconditioned stimuli those which produce, or statistically (for the species in question) would tend to produce or to be associated with, tissue injury or deprivation of a high urgency substance, such as oxygen.” “There is a family of correlated unconditioned respondents that includes pilo-erection, increased blood pressure, speeded up blood clotting time...”, etc. Or you can also do it by mentioning state α 's hookup with the

various Freudian mechanisms of defense if the students in the class know about Freud's theory of the anxiety signal in his 1926 book. By the time you have riffled through a half-dozen of these conceptual definers, everybody in the room has written down the right word (α = anxiety). I don't propose this as a touchstone in pragmatics for the adequacy of a contextual definition of theoretical terms, but something like it is an interesting exercise and, I believe, often tells us whether we are going about things in the right way. It seems to me that it is not difficult to do this for the minimally prescriptive use of the term 'rule.' We express a propositional function which has as the value of variables occurring in a certain place the actions of, say, human agents. (One might go on to specify some properties of those agents, such as that they be awake, not feeble-minded, not undergoing an epileptic psychomotor equivalent or the like; but pass that for now.) The structure and the psychological-social predicates and functors in this function are of such a character that one who is "in on the secret" (i.e., who knows we are talking about a rule) realizes that we are defining something like injustice or promise breaking or cruelty or the obligation of benevolence or whatever in terms of the non-ethical descriptors. But we don't have to say that to our Martian, or for that matter to our naive Earthling who has not been let in on the secret of what we are talking about. We are trying to see whether we can explain it to him contextually. We then explain in the metalanguage what one is going to *do* with this propositional function. We say that it is somewhat different from what one ordinarily does with propositional functions in empirical discourse—despite the fact that the expression does contain a bunch of ordinary empirical language descriptors from psychogese. I am simplifying here in talking about predicates rather than functors, although I really intend, as above, that certain obligation vectors are expressed as functors, whose value depends (as in economics) upon the values of certain other functors that *are* completely expressible in psychogese. Now what we explain to our amoral non-pragmatist Martin pure cognizer, or our naive Earthling from whom we would like to elicit the recognition "Aha! You must be talking about rules," is that the empirical values in psychogese that are associated with a human act are plugged into this propositional function and then we examine whether, when they are thus plugged in, the propositional function has yielded a proposition that is true. If so, we tell our Martian or our naive Earthling, we then do something rather special, unlike what we are accustomed to doing in physics or biology. What we do is to predicate of the particular (space time dated) act by the particular individual actor an *action property* which we label λ . We don't let our Earthling in on the secret yet, which is that λ is a notation chosen because 'l' is the first letter of the moral-legal term "licit." On the other hand, if it turns out that we get a false proposition when we substitute the actual predicates and functors of the particular action in its situation, we characterize that action by that person in that situation as β . Here we don't tip our hand to the fact that the β notation is chosen for a "baddie," that is to say, the action was illicit. Thus we explain in the metalanguage that we have a bunch of strange propositional functions that have as their variables certain kinds of empirical descriptors that are ascertained by examining the actions of a person in

a situation having certain attributes, and that, rather than doing what one normally does with propositional functions in the sciences (where we ask whether plugging in the values of the variables makes false propositions so we will junk them), here we see whether we get a false proposition *but we do not junk the propositional function* or the proposition that we get by quantifying the propositional function with a horseshoe in it; rather we use it for the sorting of actions into two boxes, the α box and the β box. Just as utopian psychophysiology could be understood by a blind physicist (except for the cognitive edge discussed in my compleat autocerebroscope paper), so a sufficiently bright and learned amoral Martian could learn a great deal about the Earthlings' psychological and sociological pragmatics in relation to the action properties λ and β thus contextually defined. He would, for instance, know that humans have some kind of a psychological impulsion to perform λ acts and that this is sometimes strong enough to countervail impulsions to β acts that are strongly drive connected (hunger, avarice, lust, rage). He would learn that very strong dispositions to β acts are often associated with the state of anxiety, and that human beings sometimes refer to how these things make them feel as "guilty," "ashamed," "frightened." He would learn that there is a set of loosely related aversive states belonging to the shame, guilt, fear of punishment, fear of disapproval cluster that seem to be markedly reduced, at least on the average, by certain *other* kinds of behaviors that human beings describe as "making reparation," "apologizing," "receiving (human or divine) forgiveness," and the like. He would learn that people with a certain brain wave configuration and with a sociopathic (49') MMPI profile show a considerably attenuated set of dispositions in this guilt-shame-reparation-forgiveness cluster, as well as a lower degree of countervailing forces provided by the individual's acquisition of a rule network. And so on. It seems to me clear that the completely amoral non-pragmatist Martian, whose vocabulary nevertheless permits of describing and causally explaining human actions and states of mind (construed in terms of the psychology of the other one), would find himself, after learning all of this, perfectly capable of saying, "These Earthlings have some special class of propositional functions that play a different rôle from the ones that they refer to as 'laws of Nature' or 'empirical generalizations.' These other kinds are what they refer to as 'rules' or as 'laws' in another sense, similar to the laws promulgated by their political rulers. Then they have characteristic kinds of conative-affective dispositions, states, and events that are connected with these in rather special ways, differing from but overlapping with the kinds of conative-affective dispositions and events that are associated with non-rules. And the most important thing about the things they call "rules" is that Earthlings do not look upon these 'rules' as being refuted (destroyed, invalidated) by the occurrence of counter-instances in their own or other persons' actions, even though the rules seem to have the form of universal statements about actions; whereas they do look upon empirical laws or generalizations as being refuted by counter-instances, provided these counter-instances are accepted by them as being genuine rather than only apparent or spurious counter-instances." Our amoral non-pragmatic Martian would understand a very great deal about what a rule is for an

Earthling and how it functions in mind and society, and his position in this respect would be very like that of the utopian psychophysicist who is blind but who knows all about the color nomologicals, including the psycho-linguistics of color talk by seeing persons.

But now the \$64,000 question recurs, to wit, is there anything that Feigl and Meehl as egos understand about moral rules that the amoral, non-acting, non-suffering Martian pure cognizer, who speaks only indicatives, would not be capable of knowing? Now I suggest to you that this situation is extremely close to that of the blind utopian psychophysicist. The only thing that the Martian lacks is a knowledge of what it feels like to be guilty, or to be impelled by one's sense of obligation to the moral law. Everything else he understands.

§ 28. There is something odd—I do not say strictly contradictory, although maybe even that could be shown by spelling out the implicit methodological commitments involved—about the logical positivist position in the 1920s concerning the noncognitive or, even stronger, “meaningless” character of ethical statements. Operating with a strict verificationist meaning criterion, Vienna positivism was in bad trouble. Hence we saw the abandonment of the strict verificationist meaning criterion in favor of progressively weaker ones, and then finally the loss of interest on almost everybody's part in even trying to *formulate* a meaning criterion, because every attempt either (a) excluded portions of science and mathematics we wanted to keep or (b) failed to exclude the metaphysics and theology that it was desired to liquidate, both failures usually being present at once! Carnap's last heroic effort in Volume I of the *Minnesota Studies* is obviously a retreat from a propositional meaning criterion to a souped up concept meaning criterion, and it is sufficiently weak that one wonders why he thought it worthwhile to bother stating it in such logician's detail. I would say that the Carnap 1956 version is, despite a faint squeak of positivist negativism still discernible, hardly distinguishable from Bertrand Russell's famous remark against the positivists (I forget just where but I'm sure you know it), “If I know, roughly, what the words in a sentence mean, and I understand the grammar of the language in which the sentence is uttered, and the sentence is well formed, then, I maintain, I know, roughly, what the sentence means—regardless of whether I can imagine how to test it.” But I digress. Let us consider the *Aufbau* program, which attempted to introduce all the theoretical entities by a reduction to phenomenal predicates. Carnap subsequently bought Neurath's physicalism—mainly, if I remember the account, because he got tired of being accused by metaphysicians and others of being a “phenomenalist” or even conceivably a funny kind of “idealist.” But the interesting thing here is that during the heyday of *Aufbau* thinking in Vienna, and the year after Carnap published the *Aufbau* [1928] and the year the crew collectively published the manifesto [1929], why should it have been said that ethical sentences were cognitively meaningless? If we stay out of the difficult and unpleasant task of talking clearly in the metalanguage about the ontological status of the theoretical entities (which was partly what the *Aufbau* and related Vienna positions were supposed to help avoid, right?), then the empirical meaning of the theoretical terms, being statements that appear “higher-up”

in the conceptual hierarchy, is all presumably given by some mixture of implicit definition (*à la* the formalist program in metamathematics) together with some “upward seepage” from the immediately experienced colors, sounds, and the like which were Carnap’s givens in his logical construction of the world. I suggest that the positivists’ real objection—I do not mean to argue *ad hominem* but merely to make an interesting historical point for your delectation—against ethical predicates, especially if we only require one distinctively ethical primitive (the others being definable in terms of “empirical” descriptors like reward, pain, promising, etc.), was a deep metaphysical feeling, which you and I still have and which this memo will undoubtedly fail to cure although it might alleviate it a little bit, to wit, that there isn’t anything “objectively out there” to which the distinctively ethical predicates or functors can “correspond.” I don’t mean to suggest that this argument was *made*. We know that the argument that was made by the Viennese was an argument based upon verifiability notions. What I am saying is that the philosophical impulse underlying the search for a verifiability criterion, which persisted long enough to go through several years of a degenerating research program (best account I’ve seen, in Brand Blanshard’s *Reason and Analysis*, Chapter V) in modified logical positivism and scientific empiricism, was a problem about the conceptual status of these strange ethical entities or relations which the positivist mind with its strong Victorian materialist science and anti-metaphysical and anti-religious bent found so puzzling. Yet it is far from obvious that the relation between theoretical entities of physics and chemistry purportedly “reducible” to phenomenal language in the *Aufbau* reconstruction could have differed all that much from the relation that would exist, in a reasonably articulate and precisified axiology, between the distinctively ethical predicates and the Prichardian intuitions *in concreto*. To carry out the *Aufbau* program Carnap had to do a good deal of over-simplifying about observational predicates and the status of phenomenal events and ordinary thing language predicates, which was of course excusable in pursuing such a research program; but I want to argue that *if* you fellows had allowed a similar looseness about Prichardian intuitions, *then* the relation between causal theoretical entities and “experiences” would not look different, qualitatively speaking, from the relationship between ethical theoretical statements and Prichardian ethical intuitions *in concreto*, would it? That is, if one gets rid of the metaphysical instinct involved in positivism, which is (as you have often pointed out) essentially “Victorian materialist scientism” in spirit, and asks how the theoretical entities in the *Aufbau* program are related to anything “empirical,” he is told that they are related by the theoretical statements to people’s experiences, that is, to phenomenal colors, sounds, touches and the like; and a sufficiently ingenious “ethical Carnap” could, I should think, have done a tolerably good job with a moral *Aufbau*, in which the “observation basis” was our ethical intuitions *in concreto*. If asked by a crypto-materialist but officially positivist critic, “But just *what* and *where* are these ‘obligations’ you speak of, these ‘oughts’ and ‘may nots,’ in the real world?”, Carnap could have replied, *Aufbau* style, “What kind of metaphysics are you trying to get us into by such questions? I do not discuss, even in metalanguage,

spooky issues about reality. I construct the world. And, just as I construct the periodic table and the ‘objective entities’ (if you must talk that way) it is about by suitable logical operations performed on expressions tokened about my experienced sours and blue-to-pink litmus elements, so I construct the moral law and the distinctively ethical relations in it by appropriate logical operations performed on expressions tokened about my experienced Prichardian intuitions *in concreto*. What more do you want of me? I suspect you of some dark metaphysical hankerings, some still uncured disease about pseudo-problems, like ‘reality’.”

§ 29. It is cheap and easy for me to say, “The main trouble with meta-ethics is the lack of an adequate object language ethics to work over in our meta-reasoning. We need a formalism in which can be stated our ‘pure postulates,’ so we can derive rich and powerful theorems. This formalism must do justice to the complex character of real moral dilemmas, to the fact that moral reasoning involves trade-offs, probabilities, means-end hypotheses, preference rules, lexical orderings (if it does), morally indifferent choices, tolerance ranges, and—especially important because so lightly treated in most meta-ethical disputes—a *quantitative* feature which the common man regularly takes for granted (‘such-and-such acts are bad, but not as bad as so-and-so,’ ‘it is important to keep promises, but not as important as saving a drowning man’).” I am not competent to construct (or select) a formalism that I would seriously propose as adequate for the kind of axiomatized and numerified axiometrics a sophisticated meta-ethical work must have available as an object language ethic fully worthy of philosophical *and empirical* examination. I suspect that the substantive disciplines which should be present “in one head” to undertake such a task would include ethics, meta-ethics, econometrics, jurisprudence, selected portions of the other empirical social sciences, and sufficient general logic and mathematics to be inventive about aspects of the formalism which may not be immediately available in econometrics and decision theory as they now stand. Furthermore, it is likely that some of the questions that will arise in evaluating the merits of a proposed formalism would require empirical knowledge about mind and society that does not yet exist; although perhaps we may safely assume that here, as in other cognitive enterprises, reflection upon the difficulties presented by a provisional formalism would itself help to guide empirical investigations in psychology and the other social sciences (including “institutional” disciplines) by making somewhat clearer what it is that we need to know about human interpersonal behavior that we do not now know.

Despite our terrible ignorance about the psychological and social facts and laws, and despite my mathematical limitations, I propose in this section to present a crude, primitive, and grossly inadequate formal structure which, if nothing else, has the merit of showing how feckless it is for us to conduct ethical and metaethical discussions in ordinary language. It is easy to see several respects in which the formalism I will present is so over-simplified that it could not conceivably serve for any but the easiest (and dullest) problems in ethical decision making. The point is not that this simplistic set of suggestions is anywhere near adequate, or even that it is the best way to start—the point

is that its weaknesses *per excessus et defectum*, not to mention the acute problems of applicational vagueness, argue strongly for my case that we are trying to do meta-ethics today as if we tried to do modern philosophy of science with Hume's examples of "scientific laws" as our object language subject matter. In making these disclaimers, I do not exclude the possibility that *something* along these general lines, at least with respect to the kinds of axiological and empirical matrices that would enter into the process, might serve as a starting point for a more competent econometric treatment.

I begin by a simplifying assumption, which I cheerfully admit is false if taken literally, that we can sort the conceptual variables into three classes. The first class of variables characterizes the properties of a moral agent's response, the response (or, better, the response class) being described in a "morally neutral way" and as close to pure behavior as possible. I shall not, however, assume that the only dimensions that can occur in such a response characterization are the physically simple ones familiar in the study of animal behavior (such as amplitude, latency, or other topographic effector-event properties). We may allow some leeway to the psychologist in the extent to which more complex conjunctions and disjunctions of effector event classes, including so called "achievement-properties" (e.g., getting the lever down, buying 10 shares of stock, turning right at a maze choice point), can be included. I do not have much to say about this that goes beyond what MacCorquodale and I set forth in our chapter on Tolman in the Koch, Estes, *et al. Modern Learning Theory* (1954) volume, especially in our section entitled "Excursus: The Response Concept." The important thing for present purposes is that, while we permit fairly complicated behavioral characterizations, including achievement-class characterizations, we are careful not to include what I shall refer to as "morally relevant consequence properties," these being assigned to my second class of variables.

It goes without saying that there is a difficult gray region here, especially as regards the matter of achievement properties. The gray region problem is particularly bad when an achievement-class property involves interpersonal behaviors. A response class defined by the effect of the organism's (perhaps effector-uncharacterized) achievement class (instead of by the effect on a manipulandum, of which pulling a chain or pressing a lever or even "using a tool" would be obvious examples from psychological research) involves human speech or gesture directed at a social object. If we say, "Jones spanked his child," this is not a response characterized in terms of effector activities except very loosely; yet it is not too difficult to decide that, like pressing a lever or turning right in a maze or lighting a cigarette, its relevant achievement class dimensions can be sorted without undue arbitrariness into the first category. Whereas if we say, "Jones spoke harshly to his butler," or worse, "Jones sarcastically criticized his wife for her behavior as hostess," and worse still, "Jones patronized Fisbee in the intellectual discussion, by treating him as if his arguments were pseudo-scientific," we are in serious trouble about the classification. All these difficulties I assume away for purposes of the present analysis, while admitting their grave implications for the task of ethical formalization.

For the sake of generality, I assume that the response properties are typically quantitative or dimensional; but it goes without saying that for some of the response properties the “quantification” might collapse to a degenerate case of an identifiable predicate being present or absent, this alternative being represented by a response dimension which takes only the two values 0 (property absent) or 1 (property present). If we spoke in terms of predicates rather than functors, we would be unable to express part of the information, whereas a general treatment in terms of response dimension functors allows us to include those properties which are of a dichotomous, trichotomous, or other nonquantitative kind as well as dimensional ones.

A given occurrence—that is, a particular response by a moral agent at a certain time and place—is then located in the hyperspace defined by the entire set of response dimensions, and the values of these dimensions define a *response vector*. A set of “similar” response-occasions will then be represented by a sheaf of nearly coincident response vectors in a response space. To avoid confusion, it should perhaps be mentioned here that not all of the effector activities that occur during a specified interval of the behavior flux (a slice or chunk out of the narrative of a moral agent’s life history) are included in the characterization of the response, for a discussion of which point see the Tolman chapter excursus cited *supra*. So we represent a response by a row vector

$$R = [r_{11} \ r_{12} \ \dots \ r_{1m}] \quad [1]$$

We do not presuppose here that all m of the response dimensions whose values appear in this row vector are relevant to all of the morally significant dimensions of the responses consequences, but we put them all in without loss of generality, knowing that the matrix formalism will take care of the fact that for *some* of the morally relevant response consequence dimensions, *some* of the response dimensions (differing for different consequences) will be without material effect.

We then assume that these m response dimensions are empirically associated, given the laws of physiology, psychology, and the other social sciences (and the state of the society in which the moral agent lives and acts), with varying probable impacts upon others. A given response dimension may have a small impact upon, say, air pollution and a large impact upon whether a second cousin’s Republican feelings are hurt. We then make the gross over-simplification of assuming that the manner in which the response dimensions influence what I am calling the “morally relevant consequences” is expressible by a linear composite, that is, that the expected value of a morally relevant consequence dimension can be written as a linear function (with appropriate weights based upon empirical probabilities) of the m response dimensions. This means that each of the n morally relevant consequence dimensions must be evaluated by multiplying each of the m response dimensions by an appropriate weight based upon empirical laws (nomological or stochastic). The natural way to represent such a state of affairs is by an $m \times n$ matrix each of whose n columns contain m empirically based impact weights showing how each of the m response dimensions influences each of the n morally relevant consequence

dimensions, there being as many columns in that matrix as there are morally relevant consequence dimensions. I shall refer to this as the “empirical influence matrix.” Its general element is (b_{ij}) where this element quantifies the expected impact of response dimension r_{1i} upon morally significant consequence dimension c_{ij} . This empirical influence matrix is then

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1n} \\ b_{21} & & & & \\ \vdots & & & & \\ b_{m1} & b_{m2} & b_{m3} & \dots & b_{mn} \end{bmatrix} \quad [2]$$

We premultiply the response matrix R by the empirical influence matrix B which yields a row vector C , the response consequence matrix, thus:

$$C = RB = [c_{11} \ c_{12} \ c_{13} \ \dots \ c_{1n}] \quad [3]$$

This product matrix C is a row vector whose general element c_{1j} represents the expected influence of a response having the response dimensions represented in the response matrix R upon the morally significant consequence dimension j , and the vector C characterizes the morally significant impact-properties of the response.

I emphasize that we have thus far not introduced any distinctively axiological functors, although we have identified certain features of the probable response consequences as being “morally significant.” As indicated in Section 9 *supra*, I conjecture that only one distinctively ethical primitive is required in axiology, to wit, the notion of obligation. The axiometric values we are going to introduce in the next step are contextually defined with reference to this primitive notion of obligation. The notion is “explained” in the embedding text and is not explicitly defined within the formalism, although the rôle it plays in the formalism, when taken together with the embedding text, “makes clear” (as clear as need be in ethics, since it is the same as we do in science!) what the story is all about. If someone wishes to describe this situation as “implicit definition” I have no objection, provided the phrase is being used to cover the integrated composite of meaning-specifiers that includes (a) rôle in the formalism, (b) Ramseyfied “upward seepage” from the collection of Prichardian intuitions *in concreto*, and (c) additional embedding text of an explanatory nature. That we construct the formalism and identify the significant variables with an eye to our values in relationship to the psychological and social empirical facts does not, of course, mean that we have to make use of these background overarching purposes, intentions and informal understandings in the guise of explicit definitions within the formalism itself, nor in the embedding text although some ethical writers seem to think so. The n consequence dimensions, while *selected* because of their moral significance, are nevertheless *defined* (in a suitably broad sense of that word, allowing the kinds of definitions implicitly, contextually, in use, etc., which we regularly find in the sciences) by psychosocial predicates and functors. The

distinctively moral term “ought” and its associated axiometric unit “ob” do not appear anywhere in that set of n consequence dimensions. The distinction is between (1) the first category of variables, where we describe a response semi-behavioristically, in terms of its effector properties or as an achievement class tied rather closely to social and biological impacts, and (2) the morally significant properties of the empirically expectable result, involving such notions as “amount of suffering produced,” “number of persons whose reasonable expectations will be disappointed,” “loss of personal dignity by the social object of one’s racist remark,” and the like. The point is that these consequence dimensions are presumed to be definable (or reducible, or somehow explainable) by reference to the facts of mind and society and without the use of the distinctively ethical predicate ‘ought’ (or its quantitative form as an obligation vector in the formalism, which we haven’t yet reached in the development). They are, strictly speaking, psychological-sociological variables, even though our selection of them for representation in the consequence matrix C is ethically motivated.

The next step is the first in which the formalism expresses magnitudes assigned by “pure postulate” to the morally significant consequence dimensions c_{11} , c_{12} , c_{13} ... c_{1n} . We represent these fundamental values by “ethical weights” in a value matrix which is an n -element column matrix V , thus

$$V = \begin{bmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{bmatrix} \quad [4]$$

On one (simplistic) view, this basic value matrix is post multiplied by the consequence matrix C to yield a degenerate 1×1 “obligation vector” = O as follows

$$O = CV = [o_{11}] \quad [5]$$

Taken by itself the numerical value of this single product matrix element o_{11} does not yield an ethical decision. It is merely a moral characterization of the response R with whose response dimensions we commenced our sequence of matrix operations, and our moral agent would have theoretically to compute a similar obligation vector for all of the possible responses currently open to him. I should emphasize that we are not assuming that the morally significant consequence dimensions are all “utilitarian,” and that the various non-utilitarian aspects stressed by deontologists (e.g., that a reasonable expectation in reliance upon a promise is morally significant whether or not the promisee “hurts” from this disappointment, although his expected mental suffering might appropriately occur elsewhere in the matrix with an appropriate primary value attached to it). A moral “optimizer,” or one might say “a rigorist,” would hold that an agent is obliged to canvass all of the possible responses available to him under the circumstances, and, after going through these matrix operations to arrive at an estimated obligation vector for each

response, make that response whose obligations vector is largest. I do not assert this, partly because I couldn't prove it but mainly because I don't believe it (and, as argued in Section 24 *supra*, I don't think anyone else really believes it either).

I think many people would object to the idea of collapsing all our information and values into one final net obligation vector [5], even if a strict maximizing rule were not imposed for choice among actions. The available actions are represented by a sheaf of obligation vectors, and the maximizer demands that we select the longest vector in the sheaf and act accordingly. I read the utilitarians as advocating this, and as believing that all the vectors are in a 1-space (dimension: total utility). A non-utilitarian can of course be a maximizer, but his vectors are in a space of more than one dimension. A non-utilitarian who accepts lexical ordering, as Rawls calls it, would not want to drop all the information contained in the consequence matrix C . Such a moralist would demarcate forbidden regions in the C -hyperspace, thereby eliminating some of the vectors in the ob-sheaf. Then he might choose to be a maximizer, or a quasi-maximizer, or a "satisficer," as regards the ob-vectors remaining. These are the kinds of issues which I think can be better considered when we have an axiomatized and numerified econometricized axiology to work with.

§ 30. In struggling with the Big Question—"What is the ontological status of the distinctively ethical predicates or functors?"—things are bad enough at best for anyone trying to reconstruct post-positivist ethics without making them worse by accepting the burden to answer questions that need not be answered, and which would not be asked in serious discussions of non-ethical domains. The Big Question may be interpreted in several ways, two of which I propose to take with deadly seriousness, recognizing (a) that they are probably not satisfactorily answerable on my analysis—I fear on anybody's analysis—(b) that they are legitimate questions and (c) of central importance, so that an ultimate failure to answer them will be fatal for any meta-ethical theory (except the unsatisfactory "pure commitment" or "Boo-Hoorah" theory). But we ought not to make the Big Question harder than it unavoidably is, so I want to clear away a couple of interpretations of it that give rise to road blocks.

First, it seems that sometimes the request for a metalinguistic account of the ontological status of distinctively ethical entities conflates what I have called "morally relevant response consequences" with "axiological functors." That is, the questioner fails to distinguish the designata of the embedding text (which explains the referents of the row vector C obtained by multiplying the response dimension row vector R by the conditional empirical matrix B) with the designata of the axiological weights in the "primary value" column vector V by which the preceding matrix $C = RB$ is to be postmultiplied. I have tried to indicate, admittedly adumbrating future social science developments in response classification and utility measurement, why I do not see any distinctively "axiological" problem as to the ontological status of the entities referred to in the course of ethical conversation *prior to our providing a textual interpretation of the last step in the formalism*, to wit, the column vector V representing our primary value commitments. To

say that the consequence dimensions appearing in the formalism as components of the product matrix $C = RB$ are chosen because they are the “morally relevant consequences” is *not* to say that they are themselves axiological predicates or functors, which they are not. The “things” involved (i.e., the *relata* of the complicated relations characterized by this vector) are, to repeat (perhaps tiresomely), easily identified entities in spacetime, to wit, moral agents (for our purposes, human beings), and sentient creatures (other human beings as well as animals or, if they are capable of suffering and of having any “rights,” plants). The problem in discussing the morally relevant properties of the response consequence vector C is one of complicated fact gathering, complicated formalism and, ultimately, complicated theory, but it is not a problem about any unique ontological status. Neither the agents or patients, neither the responses of the former nor the consequences of these responses experienced by the latter, nor the statistical relation between the two, present problems different in principle from those of empirical political science, economics, psychophysiology, reinforcement learning theory, and the like. So the Big Question must be interpreted as involving the ontological status of the designata of the interpretive text dealing with the “value” column vector V , plus the final ethical prescription that provides a decision rule.

This leads directly to the second avoidable roadblock, namely, that one seductive interpretation of the Big Question—“Just what is the ontological status of the distinctively ethical functors?”—is: “What, and where, are the entities here spoken of?” This is then further read as a demand (roughly) to the effect: “You are required to provide a metalinguistic account of what and where the ethical *relations* are (over and above the moral agents and sentient beings that constitute the *relata*); and, of course, in this metalinguistic account you must not employ any distinctively ethical object language terms, nor any synonyms, antonyms, or correlatives thereof.” In other words, the critic wants us to tell him in our semantic and pragmatic metalanguage “about the metaphysical status” of the ethical predicates, *without using any ethical predicates in our metalinguistic discourse*. This demand can be quickly and decisively rejected by pointing out that the critic has not shown—and I cannot imagine how he could show—that such a demand is (a) fulfillable for any domain of discourse, or (b) necessary as a precondition for admitting the validity or meaningfulness of discourse; and, further, (c) that he would not think, even for a moment, of demanding it with respect to empirical science or formal disciplines like mathematics or logic. So far as I am aware, there is no theorem of logic, semiotic or epistemology which reads: “In order for the use of an object language L to be licit, the user must be able to explain in his metalanguage the semantics of L without employing any terms of L or any terms of an object language L' that is translatable into L .” Where on earth would such a theorem come from?

Suppose I have been giving somebody an account of chemical valences in terms of the quantum theory of the chemical bond. And suppose when I get through my account the critic says, “Well, that’s very interesting indeed, and I think I understand all this business about completing the electron shells and sharing electrons and binding energies

and so on; and I even semi-understand Linus Pauling's resonance theory, although that's a little spooky. But what I would like you to explain to me clearly is, *what* and *where* these entities *are* that chemistry talks about?" One would reply, "I thought I had just finished explaining this. The atoms I spoke about are tiny particles which compose the macro-observable fluid which is in the test tubes that are on my laboratory desk, and these are external to my skin, and I can see them via light waves impinging on my retina, and... blah, blah, blah." But suppose the critic then says, "Oh, well, but now you are merely *repeating* yourself, telling me the chemist's story over again. I wanted you to give me a metaphysical account of the reality status of these chemical entities." Then I repeat to him that I am a scientific realist; I don't include, in my metalinguistic discourse, a fictionist or instrumentalist thesis, although I don't really need to include a realist thesis either, because I *can* simply say (in the metalanguage) that the sentences in the object language are true, and that I adopt Tarski's definition of truth and its associated (I do not say entailed) "metaphysic" in the form of the correspondence theory, suitably sophisticatedly construed. Nobody has ever held that metalinguistic discourse may not include object language sentences, and of course it generally does. Questions about the entities of chemical theory, about their "where" and "how" and "what," their "nature" and their "causal relation to my perceptions," are all questions askable and answerable in the scientific object language including, of course, the causal theory of perception. If somebody puts these "scientific" questions to me "philosophically" by asking me the same thing over again, but this time in the metalanguage (asking, for instance, not "How big is a hydrogen molecule," but, "What is the bigness of a hydrogen molecule, and where is it?"), he cannot estop me from using words like "hydrogen," "molecule," "distance," "size," "angle," "mass," by telling me that he intends now to put a philosophical question in the metalanguage about metaphysical status. The point is that I cannot discourse about the subject matter of chemistry, even about my epistemological and ontological views concerning the entities with which that subject matter deals, if I am arbitrarily forbidden to use any chemical words in the process of answering questions. Somehow one of the tempting misapplications of Hume's Guillotine has come to be that one must be able to give a metalanguage account of his moral metaphysic without using any distinctively moral terms. Until somebody can show why this should be considered necessary in ethics when it is not considered necessary in any other cognitive domain, the best reply to this particular interpretation of the Big Question is: "I can't do it, but I am under no obligation to do it; and you can't do it about any other subject matter that you discuss either."

§ 31. It could plausibly be argued that there is in ordinary language some built-in prescriptive component to such expressions as 'equal treatment,' 'justice,' 'fair return on a risk,' and the like, and I have no reason for disputing this as a statement in the descriptive semantics of ethical discourse. But it seems to me that we have here a circumstance analogous to others involving a Carnapian explication, that analysis reveals separable components in the descriptive semantics of a certain kind of expression. If one perennial source of dispute or misunderstanding in the domain has been the presence of

such components (intertwined despite their distinguishability), we do not commit a needless violence upon ordinary usage to propose, pending constructive criticism of the explication so far as we can carry it at a given time, a revised meta-language semantics that separates the components of meaning without thereby prejudging anything as to their synthetic (postulational or empirical) connections. It is as if I should be forced by legal, commercial or other “common life” conflicts to reflect upon the meanings of a common language expression designating, say, a familiar article of furniture or a human interpersonal action (e.g., “buy a credenza,” “challenge to a duel,” “make a plea”) and, upon such reflection, pursued in some philosophical depth (despite the ordinariness of human situations out of which the necessity for such analysis has arisen in the conduct of life), I find that a common expression such as ‘settee’ or ‘grandfather clock’ involves in its semantics a vague mix of *manufacturer’s intentions plus commercial specification plus retailers’ verbal habits plus homeowners’ intentions plus materials plus physical dimensions plus ordinary use of the object* and the like. In settling a particular lawsuit about furniture it is imperative that these components be distinguished and, perhaps, that some of them be considered definitive rather than others. The fact that everyday discourse involves a simultaneous reference to distinguishable properties and dimensions that we ordinarily do not have occasion to split apart is not a justification for continuing post-analytically to use the word ‘settee’ or ‘contract’ as if we had not previously carried out our linguistic analysis. Not being a disciple of the later Wittgenstein, I do not begin with a presumption in favor of ordinary language or an expectation that whatever approximate preliminary semantic rule might be proposed as an explication of the crude language of a tribal society or of the contemporary Western citizen will be adequate to our needs. *A fortiori*, I do not impose conformity to vulgar speech as a requirement. I assume with Feigl and Maxwell that ordinary language needs reforming more often than not, the more so in a field such as meta-ethics which has remained intensely controversial and conceptually muddled for centuries, and which has received such a relatively small share of clarification under the impact of modern analytical philosophy beginning three-quarters of a century ago.

Hence if someone says that no analysis of a characteristically juridical or ethical term, such as ‘justice,’ which omits its distinctively prescriptive (judgmental, evaluative, obligatory, imperative, praise or blame) component can be accepted as an adequate explication, my reply would be a strong denial that such a criterion based on the facts of ordinary descriptive pragmatics should be accepted as a touchstone or criterion in efforts to clarify meta-ethics, any more than a subjective feeling of push or pull, or before and after, or intrinsic necessariness should be absolutely required of a post-Humean explication of the concept of causality. Analogy: Even those who (like myself) wish to preserve some notion of necessity in our ideas of causation (and social science, I think, can hardly settle for anything that doesn’t include such an element) would not require that an analysis of causality should contain this element as somehow conceptually welded to the other elements (e.g., regularity or predictability) that may emerge upon careful analysis.

The fact that I want some notion of necessity other than mere correlation to be preserved somewhere in an analysis of nomologicals does not mean that I have therefore to reject a complex and sophisticated analysis on the grounds that some component of it is reconstructed as a partial account and that that component leaves out my pre-Humean feeling of necessary connection. If it turns out that part of the disagreement arises from such a welding of two separable components, separating them analytically is all the more important.

Such a “separatist” view shows its conceptual power when we turn to the standard anti-emotivist objection which has been advanced, I suspect by every variety of anti-positivist or anti-emotivist during the last 75 years (Moore, Blanshard, Ross) and which we find earlier in Sidgwick, and still earlier in Price and Butler, namely, that if a pro/con attitude or an imperative on the part of the speaker or an alleged introspection of one’s own sentiments or any of the other variants of what in modern times is often disposed of by calling it the “Boo-Hoorah” theory of ethics were a correct account, then obviously two informed rational reflective intellects could not carry on an extended argument over matters ethical. And since we all know that such persons *do* carry on such extended arguments, we know immediately that the Boo-Hoorah theory is radically false. That this cannot serve as a definitive rebuttal against a sophisticated form of the Boo-Hoorah theory is seen by realizing that there are at least a half-dozen distinctively cognitive components in ethical argument. Some of these are empirical about one’s own introspection; some are empirical about other persons’ moral reactions; some are empirical about means to ends; some are empirical about changeable versus relatively unmodifiable basic features of human nature, or of the human condition as *homo economicus* and *homo politicus*; others are formal or logical, involving questions of subsumability, completeness, contradictoriness, or derivability. It is not a sound objection to an imperativist or emotivist theory concerning rock bottom ethical postulates that people of good intelligence and good will can argue and can persist in disagreement, just as it would not be an objection to, say, Popper’s theory of the unarguability of protocols that scientists disagree about whether a certain protocol should be admitted, or an objection to a postulational view of mathematics that mathematicians disagree about the validity of a given proof or about the three competing metamathematical theories as to the very nature of mathematical truth. I can put it simply thus: It cannot be an objection to any underlying metatheory of ethics that people have ethical disagreements and carry out complex cognitive discourse about their ethical differences, so long as a reasonably adequate account of how such differences can arise and how such arguments can be cognitively significant is available to us, given the meta-ethical theory in question; and on almost any theory (any that has been respectably defended—the emotivist theory in particular, but others as well), we already can see how it could be possible for such extended disagreement and genuinely cognitive discussion to arise. Let the rock-bottom ethical postulates be characterized any way you like, epistemologically and as to their ontological status and their grammatical character; all remaining sources of inference and collision have already been

set out. Consequently, the fact of dispute and persisting disagreement and the difficulty of conceptual analysis or application to moral cases has been explained in general and in advance. Each meta-ethical theory is entitled to begin with this list of a half-dozen ways to disagree (and the possibilities for cognitive reconstruction), and in that respect one is in no worse shape than the others.

Thus one need not dispute whether words like ‘unfair,’ ‘cruel,’ ‘dishonest,’ always, generally, sometimes, rarely, or never have a judgmental, praise or blame, prescriptive, obligatory, or imperative component attached to them, whether in ordinary language, in law courts, or in a seminar on jurisprudence or ethical theory. I myself am inclined to think that they almost always do, in almost all contexts. But in an ethical theory or jurisprudence seminar I would want to separate the descriptive from the prescriptive component. If anyone finds this does too much violence to his linguistic habits, then I would be happy to have a conventional semantics of the words ‘just,’ ‘unfair’ and ‘cruel’ introduced by recourse to Korzybskian subscripts, or by the usual dredging up from the “dead” languages of an antiseptic term. The point is that we do not wish to be forced to say that a word like ‘cruel’ must always inherently designate something more than, roughly, “producing suffering for the sheer pleasure in perceiving it,” that we cannot say this without involuntarily saying “and, furthermore, I don’t like that, and you shouldn’t like it, and I command you not to like it, and shame on you if you do anyway.” What stake do we have in saying that the one is inherently connected with the other, prior to the completion of our metatheoretical analysis? Why would it be an analytic advantage to have them semantically fused, rather than to keep them semantically distinct so that we can *subsequently* reflect upon whatever kind of proposition (including, if you like, a synthetic *a priori* proposition) proceeds then to link them? I do not see why anyone thinks it could help me in examining the status of my imperative, “Do not inflict suffering for the sheer fun of it” (whether it is an ethical postulate or a theorem or a Prichardian intuition *in concreto*) to pack the imperative, hortatory, or disapproval components of my discourse into part of the *definition* of ‘cruelty’ as inflicting pain for the fun of it, or the definition of a ‘sadistic person’ as one who has a tendency to be cruel. If there is some way in which it can be shown—even on a synthetic *a priori* judgment of such a strong sort that we say that one who cannot make this judgment is morally blind or desperately wicked—then surely we can arrive at that truth without packing it into the meaning. But if it is packed into the meaning as part of a two-component term, we might have grave trouble separating the issues. Of course what I want to suggest as the outcome of these reflections is that the whole gamut of distinctively juridical and ethical predicates and relations can be reduced to terms of psychologese and sociologese, except for the distinctively ethical weights in (and the text surrounding) our final post-multiplied value matrix.

This is perhaps as good a place as any to say something about my view of ordinary language, which it must be obvious to the reader I do not take as a criterion of anything, despite some scattered (and I hope legitimate) references to “what we would ordinarily say.” To lay my cards on the table by a general remark: I look upon the ordinary language

movement as one of the most useless developments that has occurred in the entire history of philosophy and one which has not only made no positive constructive contribution to any important scientific, moral or religious issues, but which has done active harm by deflecting the attention of some able young intellects to Oxbridge “puzzle-dissolving” who might have made a significant contribution had they applied their talents to something worthwhile instead. More in detail, then: I have seen no affirmative showing, either on theoretical grounds or from the history of science, that reflection on my sainted grandmother’s verbal habits, or even the verbal habits of a don, has either clarified theoretical concepts, suggested experiments, or helped provide causal or other explanations of anything. Since earth, air, fire and water have not turned out to be powerful for chemistry, nor the divine right of kings for political science, nor what the neighbors think for sociology, nor even semi-sophisticated academic usages like “quantity as such” in the analysis of such concepts as numbers, it puzzles me why anyone would expect that ordinary language would be useful in thinking about jurisprudence, ethics or psychology. It is sometimes said that we are all human beings with ideas and motives and we have all lived in the world with other people and made judgments about their conduct whether moral or merely descriptive, so our ordinary language ought to be adequate to these modest jobs. I should like to state loudly and clearly as a psychologist that I do not find that argument possessed of the faintest merit. We have all spent a lot of time in the weather, which doesn’t tend in the least to prove that the Farmer’s Almanac is suitable to conduct the science of meteorology, either as a theoretical or as a practical enterprise. We have all eaten food, and most of us have prepared it in the kitchen or elsewhere, but that does not contribute much to the science of nutrition, let alone biochemistry. To take a classic example from my own field of expertise: So far as human conduct is concerned, it is a truism of scientific psychology that the existence of a trait-name in the dictionary has only the feeblest tendency to show the empirical reality of a trait and negligible probative value as to its pervasiveness or importance either in the description or prediction—let alone the explanation—of human behavior. For a half-century, at least since Hartshorne and May (1928), it has been taken for granted by psychologists that the first thing you have to do when you talk about any kind of trait is to carry out the necessary laboratory or statistical research to show that the trait *even exists*, let alone has any powerful explanatory or theoretical value. Allport and Odbert found some 18,000 trait-names in *Webster’s New International Dictionary*, most of which are worthless for scientific psychology either in its theoretical or technological aspects. Before Linnaeus came along, people using ordinary language classified the whale with the pickerel and the bat with the chickadee, whereas a decent scientific classification puts the whale and the bat together. The ordinary language of rocks and peonies is inadequate for geology and botany. It may be argued that these examples differ from examples of human conduct in either its causal or moral aspects, because in these other cases of physics and chemistry and meteorology and the like the ordinary language used was a superficial, phenotypic, catch-as-catch-can, rule-of-thumb classification that does not go to underlying deeper, pervasive, and more

scientifically respectable causes. Whereas in the case of human conduct ordinary language should suffice because that's the level which makes a proper analysis of such things. This is, of course, a gross *petitio principii* by an ordinary language defender, since that is precisely what is at issue when we have an argument about whether ordinary language needs to be reformed or not (Feigl and Maxwell, 1961). Having now spent thirty-five years studying psychology both in the laboratory and in clinical practice, I do not have any more reason to think that the ordinary language of human motivation and cognition or of mental illness or normality is adequate to its theoretical and engineering tasks than I do to think that earth, air, fire and water are adequate to the tasks of a theoretical science of chemistry or an applied science of chemical engineering. Ordinary language is not even adequate to a purely descriptive behavioral analysis, as Skinner has shown in the field of operant behaviorism on the one hand or as we know from the psychometrician's field of mental measurement on the other. For one who (like myself) does not think that operant behaviorism or purely blind empirical test construction can approach an adequate account of the nature and workings of the human mind, the fact that even these refined *nonordinary* languages at the descriptive level go far beyond common sense discourse about human beings makes an "ordinary language adequacy postulate" about human behavior not merely unpersuasive but downright ludicrous.

Two things I would grant to a defender of ordinary language in any field of theoretical or practical knowledge, from dietetics or meteorology to the theory of dreams or jurisprudence: (a) The ordinary use of words should lead us to have a careful look at what they roughly seem to be pointing at, lest we miss something worth noticing, while being quite prepared to find that most descriptive terms do not carve nature at its joints and most explanatory terms are pseudoexplanations and superstitions; (b) It is obvious that even a refined neopositivist definition or Carnapian meaning explication ought to pay at least some heed to the lexical (dictionary maker's) empirical component, because it is pointless to use a word that is already embedded in discourse to mean something totally novel. Even this last I would qualify by suggesting that it is usually useful to drop a piece of ordinary language following a scientific clarification of a concept or a good theoretical explanation of the appearances, since it is almost never the case that ordinary language slices the pie properly from the standpoint either of administrative or predictive classification and extrapolation, let alone from the standpoint of theoretical causality. Hence the common recourse in all of the sciences to introducing technical words whose roots come from "dead languages" so that people will not carry over the meanings of common life illegitimately. But if ordinary language already has a word in it like, say, "insane" and I am dealing with a subject matter such as the insanity defense to a criminal charge where some of the purposes of ordinary language get closely mixed in with technical psychogese or psychiatrese or even behavior genetics, while it might be sensible for me to liquidate the word "insane" entirely from the forensic psychologist's vocabulary (some people have advocated this and I am quite sympathetic to it myself), it would obviously be foolish to retain the word "insane" and offer as a new meaning "any 6 legged member

of the arthropoda” or “ anybody under the age of 21 years.” Those two functions of ordinary language understanding and communicative convenience are rather slight concessions to one of the major movements in contemporary philosophy, but I am satisfied in my own mind that the movement deserves no more.

§ 32. While the matrix formalism of Section 29 (*supra*) is in one respect reassuring to the would-be ethical “objectivist,” as it points to a way in which convergence among ethically articulate and sensitive persons might be achieved, from another point of view it is disconcerting for one who wants to claim some kind of objective, “out-there” truth or validity for his primary axiologicals, because one has a hard time imagining just how anything outside the human mind could “correspond to” a *specific numerical value* in the final post-multiplied value matrix V . Thus, if we think that a blindingly obvious qualitative postulate (e.g., “You have a *prima facie* obligation to keep a promise,” or “Cruelty is wrong”) is objectively true, that it reflects something metaphysically correct about the relations between a moral agent and another sentient being (which would be true even if all moral agents became sociopaths overnight, or if we all had hopelessly corrupt ethical upbringings), it seems rather strange if nevertheless the quantitative values in the column vector V are in some sense unarguable differences in taste, preference, or commitment. This impression of paradox is perhaps somewhat softened by reflecting on semi-quantitative postulates of the sort that Sir David Ross and others have called preference rules, rules for resolving an ethical dilemma when two *prima facie* obligations, each formulated in qualitative terms, are in conflict (e.g., “Whether a countervailing consideration can out-weigh the obligation of a promise depends partly upon the expectable severity of the promisee’s disappointment”). Maybe there is some ingenious way of going from the qualitative through the semi-qualitative preference rules to an optimizing assignment of the primary value weights in the value vector V , and I hope to do some constructive work along those lines. But, I repeat, it does seem sort of counter-intuitive to say that one could be objectively wrong about the numerical values of his personal ethical V -vector; yet it seems (to me) equally counter-intuitive to say that one cannot be objectively wrong about the V -magnitudes, although the qualitative moral features they quantify *are* “objective,” “out there,” “not mere psychological commitments.” It sounds rather like a physicist or psychologist saying that (a) color words denote objective colors, (b) color numbers quantify whatever quality it is that color words denote, yet (c) color numbers are purely subjective.

I suppose one could say that the numerical values of V are somehow “out there,” given a properly mixed correspondence-coherence theory of ethical truth, and that line should also be pursued. Second, as I mentioned in Section 17 *supra*, probably ethics doesn’t permit or *require* sharp numerical values. A metatheoretical statement of this might be acceptable despite the claim for objectivity in the first two kinds of postulates. Can this be made a coherent position, I wonder? Working out the empirical facts of human psychology and society in detail that an adequate ethical system’s application requires, together with a really thorough working out of the ethical formalism, it might

turn out that if we had all the empirical facts we would be able to narrow greatly the coherent values that can be assigned to the v 's in V , especially if reasonable coherence with the collective set of Prichardian intuitions *in concreto* is included in our requirements for an adequate ethical system.

A general question of intrinsic interest is, "What are the knowledge conditions for going from qualitative premises to quantitative conclusions?" (Remember Eddington's crazy cosmological claims?) In ethics, economics, politics, and jurisprudence the obvious way for such a transition to occur is by conjoining qualitative axiology with quantitative empirical statements (about mind and society), and I find it hard to conceive any other way to do it.

All this is hard stuff to think about but surely worth pursuing more deeply than I can here.

§ 33. Westermarck at times seems to write as though having proved the fact that expert, thoughtful people disagree *on some postulates*, therefore he has proved ethical relativity, and that a "moral sense" capable of discerning objective ethical truths cannot exist. Aside from the objections to this argument which I have offered in several sections *supra*, there is also the point that Westermarck seems to mix up the question whether people of different persuasions (or, even more in his work, members of different cultures) disagree on *some* postulates, ergo they do, or could, disagree on *all* postulates. His own data do not clearly support this even if taken at face value, as was pointed out by numerous subsequent anthropologists. If there are *some* postulates that people in different cultures—at least representative members not diagnosable on other grounds as criminals or crazies—converge upon (e.g., wanton cruelty for the pure fun of it is wrong), maybe only these on which there is high consensus are part of the minimal set of pure moral postulates. This makes me think of Euclid's parallel axiom. Is there any non-Euclidian geometry that, say, denies the transitivity of the (=) relation? (*Cf.* discussion of Hilbert's axiomatization in Blanché *Axiomatics*, where it is pointed out that Euclid's "axioms" are *general logical truths as applied to quantities*, whereas his "postulates" are specifically geometric.) Also the former seem "self-evident" (for that reason, partly?) and the latter not, or at least less so. Blanché reminds us that the notion of a point lying outside a plane can be set out with purely logical (set-theoretic) ideas, whereas "P lies outside sphere 5" cannot be handled without distinctively geometric (spatial) ideas.

From $(\exists P) (\exists y) (\exists x) (x \text{ holds } P.y \text{ denies } P.x \neq y)$ it, of course, doesn't follow that $(P) (\exists x) (\exists y) (x \text{ holds } P.y \text{ denies } P.x \neq y)$. From the fact that there is a moral principle upon which two people disagree, it does not follow that for every moral principle there are two people who disagree. Query whether anyone would ever make such an elementary mistake outside of ethics?

It occurs to me to raise the question what kinds of non-Euclidian geometry would arise if we deny other postulates than Euclid's (was it actually Playfair's?) Fifth? Do we get a "geometry" at all? Presumably we get a *something* (since the Euclid postulates are

independent, so denying one, two, three, ... will not generate contradictions), but is it a meaningful something? And if meaningful, is it a geometry? While I think worrying about the meaning of words (usually) is, as Popper says, a sure road to philosophical perdition, it is here rather interesting to reflect on the fact that theoretically (depending upon how many of Euclid's primary statements one calls postulates and whether they are redundant, matters I know little about) there should be some $2^{10} = 1024$ distinguishable geometries, that is, all of the non-Euclidian geometries obtainable by substituting for each of his 10 axioms or postulates its contradictory. I am kind of puzzled to find that the ones we always hear talked about are those obtained by denying Playfair's Fifth, saying either there is no line that can be drawn through a point parallel to a given line, or an infinite number of such. I inquired about this with some of your brethren in philosophy. John Earman points out that there is, of course, a non-denumerable set of non-Euclidian geometries; but that isn't very interesting for present purposes because it relies upon the fact that the values of the space curvature tensor can be assigned as many different numerical values as the cardinality of the continuum. But Earman also asks the more interesting question, whether even such a rock-bottom "geometrical" principle as the principle that $d(P, P) = 0$, that is, the distance between a point and itself is a nothing distance, could be denied and yet we would still call it a geometry? I'm not sure how much relevance this has to the problem about ethics, except maybe this: One might insist that we should not call an axiology an "ethic" unless it included certain postulates. Others we might allow to be modified and simply distinguish "different ethics" (as we distinguish "different geometries"). It seems to me that we could live amicably and without killing each other in a world where some people had a rule utilitarian ethic and others had an act utilitarian ethic, despite the fact that they would come to different moral decisions *on certain hard cases*, although by and large they wouldn't differ much; but we cannot live in a world with people who have, say, Hitler's ethic. I suppose there might be only a very few postulates (one? this I doubt) which we would stipulate as necessary before labeling someone's "action-deciding-system" an 'ethic.' One strong candidate is, of course, the kind of neutrality principle that forbids me to assert where R ranges over moral relations and x,y over persons,

$$(\exists R) (x,y) [(x = \text{Meehl}) \supset (xRy)][\sim(x = \text{Meehl}) \supset \sim(xRy)]$$

stated variously by Jesus, Buddha, Confucius, Kant and Co.

John Wallace, responding to the same question as Earman, says that when you go back and read Hilbert with this question in mind, you do get an impression that Hilbert believed the axioms of geometry can be split into two groups, in the sense that some axioms are up for grabs and will generate interesting geometries, whereas other axioms are *not* up for grabs if what we are axiomatizing is a geometry. A possibly related point is that historically the reason the parallel postulate is the one people chose to fiddle with (even Gauss) is that *it lacked that quality of self-evidence* which the Greeks (and all

philosophers and mathematicians until the rise of Vienna positivism) took as one of the properties of anything you should call an ‘axiom.’

§ 34. In Sections 13 and 24 *supra* I discussed the mutual “control” exerted in both directions between Prichardian intuitions *in concreto* and primary axiologicals (value postulates). I analogized this to the mutual control between facts and theories in the empirical sciences, pointing out that the sometimes alleged circularity involved here is not a real circularity, or at least not viciously so, because the mutual control is between one *collection* of statements and another *collection* of statements. These two collections are related in logically complex ways rather than in a simple linear syllogistic way of the kind that *would*, if present, constitute a *petitio principii* fallacy. But how about the relation between primary *v*-weights in the post multiplied column vector *V* and new “facts,” such as one would come upon in the course of developing an adequate ERVE in the Crosby-Meehl ERDERVE?⁵ Can the positivist critic invoke Hume’s Guillotine here? Suppose author John Howard Griffin says that his primary axiologicals as applied to the Archie Bunker case discussed in Section 19 *supra* were considerably altered by the “facts” he experienced when disguised as a Negro in his famous study reported in the prize-winning book *Black Like Me* (1961). Suppose he tells us that if we really believe what we say about the ERDERVE precondition for including persons in the group of those whose reflective Prichardian intuitions will be treated as a kind of “ethical protocol” analogous to the observation statements of the sciences, we should try dyeing our skin black for a while, or, failing that, we should at least read his book and other books about the black experience, do some crisis intervention in a black ghetto, and so forth. Does Hume’s Guillotine make his position untenable? It does sound as though he was telling us that you can in some way “derive values from facts.” There are several answers to this which need development and clarification at length, but I will merely list them here so I don’t forget them:

- a. Following my ERDERVE-seeking engagement in the Griffin Game I may speak loosely by saying, “*Therefore*, I have altered my *V*-matrix,” a locution which makes me appear subject to philosophical decapitation by Hume’s Guillotine. But perhaps I have exposed myself needlessly to this risk by the way I chose to say it? If the *V*-weights are “pure postulates,” or even if (see Section 37 *infra*) they are “absolute personal commitments,” then I have changed the *V*-matrix psychologically as a *causal consequence* of certain experiences but it is misleading for me to talk about this as though they had been, so to speak, changed “logically” in the sense of a new proposition producing a new “argument.” Who can complain (including Hume!) if, my *V*-weights being pure postulates, I adopt a different set this month from the set I held last month? Although this line is a bit disingenuous,

⁵ [“The acronym ERDERVE—Extended Rational Discussion based on Extensive Real and Vicarious Experience—was ERDEVE in the earlier version; I subsequently added the ‘Real.’” *Cf. footnote 1.—LJY*]

it is nevertheless a technically adequate reply if Hume's Guillotine is being used in its strict sense, that is, as a reminder that one cannot derive a sentence containing ethical predicates or functors from sentences which do not. The answer that a player of the Griffin Game gives to Hume in this context is that he didn't "derive" or "argue" the V-weights, he simply adopted new ones. He says, "It is you, the Humean critic, who are setting up the (invalid) connection between the Griffin Game experiences and the alterations in the postulates, and then cutting this connection by using Hume's Guillotine. But I never made it."

- b. Means-end relations are, however, clearly empirical and they, of course, are not subject to Hume's Guillotine. That is, part of what I learn in the Griffin Game is a subtle, crude, personalistic *but still "empirical"* means-end readjustment. For instance, maybe I don't need to alter a primary postulate regarding human utilities or psychic pain or whatever, but I do alter one concerning freedom and autonomy in the Elk's Club, because I get an incoherency in the system when I try conjoining it with certain means-end statements which are empirically alterable by my Griffin Game experiences and are not subject to Hume's Guillotine (because they are means-end indicative assertions about social and psychological facts).
- c. The concrete Prichardian intuitions can, of course, change as a result of my Griffin Game experiences in my search for an expanded ERDERVE basis, so that if V were fixed, and if all the facts, including the means-end facts of (b) were fixed, then we would find our new Prichardian intuitions *in concreto* would not be derivable from the old system, and consequently some readjustments are required in the values of the V-vector components.

§ 35. In fiddling with the weights (values, utilities) in the primary axiological post-multiplier matrix V, we have tentatively concluded *supra* that these weights are to be adjusted so as to approximate two things:

- a. Success in generating "right results" judged against reflective Prichardian intuitions *in concreto*.
- b. Matching our reflective intuitions *about* the primaries themselves. But here we want sensitive, articulate, rational, and "experienced" ethical judges to be the providers of our protocols for (a) and our "pure postulates" for (b). To get such people—I remind you again of the noncircular character of this by analogy with the calibration of respected experimentalists in the sciences—we want:
 - (1) Other tests of brains, rationality, information, openness, flexibility, responsiveness to objections, arguments, nondogmatism.
 - (2) ERD on the subject matter of ethics specifically.
 - (3) ERVE in the major domains of life to which the axiology is applies.

Now we draw some conclusions as to asymptotes by studying empirically how people's values move, and how the derivatives $\frac{dv}{de}$ differ with personal traits, as more ERDERVE is accumulated by individuals in selected groups. We might prefer as best estimates of the "correct primary values" those that are thus converged upon.

This sheds some further light on why it is not viciously circular (or ethnocentric or elitist) to devalue the ethical deliverances of tribal societies, which we discussed in sections *supra*. I point out that we make such comparisons as to people's overall average competence in specified domains regularly, outside the areas of ethics and esthetics. As to circularity, while *some* substantive deliverances of individuals and groups are relied on in calibrating them as to how seriously we will take their *other* substantive deliverances, it is worth nothing that the complex non-circular control of substance by persons' view of substance is not always based upon substance, although this is surely part of the picture. The calibration of persons or groups as to the weight we should give their deliverances, whether about concrete moral cases or about general moral principles, includes at least the following, only one of which is substance:

- a. Substance. A whiff of circularity, but we recall the class-to-class aspect, discussed at length above, in regard to protocols and theories in the sciences.
- b. Articulateness.
- c. Coherency (I think I mean something here a little more general than strict logical consistency, although I'm not sure).
- d. Ability to reflect rather than to merely repeat, "It is not done," or "Our Sun-people know better," or "We English think that's bad form, old boy," or "Ffaaugh, how *wicked* of you to even think such things!"
- e. Power, richness, and subtlety of the required meta-apparatus. I get the impression from anthropology that despite a great deal of complexity as to object-language distinctions in some preliterate cultures, the meta-apparatus is dreadfully impoverished; and a *meta-meta*-apparatus seems almost non-existent. As you would no doubt expect of a semi-Spenglerian like me, I see the existence of rich, complex meta-discourse as one litmus test for a "High Culture."

§ 36. Another troublesome disanalogy between an allegedly objectivist or realistic ontology of ethics and that of physics (or molecular biology, or even psychology) is the problem of what we would consider the ethical equivalent of the astronomer's Copernican Turn? That is, in physics a causal account of knowledge is (ultimately) given in the Copernican Turn. But what corresponds to this account when we look at the primary axiologicals in ethics? We don't think of an external something *causing* our Prichardian intuitions, do we? Or if we do (as, say, something like our having an event or state of sympathy or conditioned guilt feelings elicited by our seeing a person's distressed or disapproving facial expression or whatever), these surely do not play the kind of rôle

that the nomologicals of physics are presumed to play in the causal theory of perception and, ultimately, in the causal theory of epistemology.

That's a toughy and I have no reply to it at present. However, as a starter I would point out that even in physics what literally causes our perceptions are *particulars*, space-time dated events which satisfy nomologicals, not the nomological themselves. It would not really make any sense to say that, for example, the law of gravity causes my experience of bodily weight, or that Maxwell's equations cause my visual rawfeels. The particulars that cause our perceptions and thus give rise to our knowledge are several steps removed from the fundamental nomologicals of physical science, and it may be that we can get clear what is the proper ethical analog to this complicated causal situation. But I have reflected only superficially and briefly about this important problem as yet.

§ 37. Finally, lest you think I am taking pot shots at the Vienna position from different angles but trying to avoid the positivistic-relativistic-subjectivistic Big Question—"In what sense can ethics be true or false, correct or incorrect?"—let me close with this compact and somewhat dogmatic section which deals with the question, "What metaphysical boat would we be in if we were finally forced to agree with the emotivist that the primary V-matrix represents a batch of personal commitments or moral resolutions *and no more*?" That is, despite all of the cognitive components discussed above, suppose we were to reluctantly admit that the Boo-Hoorah theory, while it does not apply in its crude form to each and every ethical theorem or application thereof to a fact situation (because of the clearly cognitive features involved in those judgments and our arguments about them), nevertheless does apply to the primary axiologicals, to the primitive ethical statements which, following Donald Cary Williams' 1933 paper, we have called "pure postulates." I will make the main points I see here briefly and without proof, to be expanded later.

- a. First let us set the V-matrix weights aside and consider qualitative postulates only (see Section 32 *supra*). We hope to work out how one might arrive at V-weights from a qualitative start plus semiquantitative preference rules plus enriched knowledge of the empirical facts of mind and society plus ERDERVE. So we might here best consider a *prima facie* obligation such as "One ought not to be cruel," stated qualitatively only.
- b. What can the would-be objectivist assert that he wants to assert, even if the primaries are pure postulates and, speaking in the pragmatic metalanguage (which, I repeat, no one can *force* us to speak in!), might be described as "basic personal commitments"?
 - (1) The objectivist can of course employ the morally relevant descriptors quite apart from the V-weights, that is, he can obviously say things like "Jones is unfair," "Smith is cruel," "Fisbee breaks promises," "Brown doesn't apply the same rules of conduct to himself as he seems to expect other people to live by and becomes indignant when they don't." I remind you that we made a distinction above between morally neutral behavior descriptors and morally

relevant behavior descriptors, emphasizing that the second class of descriptors is still formulable by reference to the language of psychologese but picks out features or characteristics of morally significant actions although it does not involve using the moral values or weights or even the qualitative postulates themselves. We have agreed that I can find out whether somebody has made a promise and whether he breaks it, without having asserted the distinctively ethical propositions that one is obliged to keep his promises and that promise breaking is wrong.

- (2) Questions of derivability, consistency, the presence of ethical gaps, questions of subsumption of fact situations under a given ethical theorem are all clearly cognitive and assertable regardless of whatever decision we might reach as to the metaphysical status of primary axiologicals.
 - (3) The matter of ERDERVE convergence is investigable like any other social and psychological question and we can of course include our empirical finding that extreme non-convergers are defective *on other grounds*. Sociopaths, political fanatics, or generally muddle-headed persons (who are, therefore, muddle-headed in their thinking about ethics) are examples with which we are all familiar.
 - (4) Statements can be made about short term and long term social results in terms of the utilities and disutilities of individuals and the likely effect upon social institutions if a primary axiological A_i were denied by most people, or if most people verbally acceded to A_i but regularly acted in opposition to it and without subsequent guilt feeling or impulses to make restitution and the like.
 - (5) It is investigable whether persons tend to feel the same way when they are in the agent position versus when they are in the patient position in a particular ethical relation, so that our objectivist can say such things as, "I notice that most people object strenuously to cruelty when they are the object of it even though they may on occasion not object, or not so strenuously, e.g., when they are the agents of it or the passive spectators who permit it."
 - (6) We can investigate empirically whether the proposed ethical postulates generate theorems (which apply to fact situations) that tend statistically to map people's Prichardian intuitions, especially those of ERDERVE sophisticates.
- c. But what can our would-be objectivist assert in the metalanguage *about* his ethical postulates? I am certainly not about to try to clear up that question in a couple of paragraphs! However, I think there is an important negative point to be made in that respect, which is made by Munro in his excellent book cited *supra*. To say that the primary axiologicals have the status of pure postulates, and even further to say that the reason they have this status is that they represent some funny mixture of

Boo-Hoorah and personal action commitments and *nothing indicative* (this last italics phrase being one, I repeat, that the positivist-relativist-emotivist critic has not succeeded in proving, despite making us post-Viennese feel uncomfortable), it is *not* required to say, “Since there are pure postulates, they are not true or correct or morally sound,” or, the favorite hang-up of the strong joyous relativist, “Since my ethic is postulated and expresses my commitments, I of course must hold that yours are just as good as mine.” There isn’t any reason why a person should feel obliged to say this in the pragmatic metalanguage about other people’s ethics, and there isn’t any reason why he should say in the semantic metalanguage that his own ethical statements are “not correct.” The reflex response to this is, of course, to ask what you would do with somebody who holds different ones, but I trust I have brought about at least some attrition in the potency of that complaint.

One has (I say, mostly because of our Victorian materialism and the positivist bred in the bone belief that only stones and butterflies are “real,” “objectively there,” “existent apart from human minds,” but, I freely admit to you my old reformed positivist buddy, also a genuine philosophical problem about the ontology of ethics) an irresistible impulse to ask where in the external world outside of our human minds the ethical relations designated by the ethical predicate *exist*. I hope I have relaxed you about the entities which are the range of the variables in ethical propositional functions that become ethical propositions, namely, they are moral agents and the converse domain of ethical relations which consists of other moral agents and (to cover the case of animals) “other sentient beings.” So it’s not the ontology of the subjects of ethical discourse that is troublesome, they are people and infrahuman animals capable of suffering. It’s the obligation relation itself, it’s the whatever that is expressed by our primitive ethical predicate ‘ought,’ that’s the tough one to ontologize about.

But one must be careful not to conflate the epistemological with the ontological problem, despite their intimate connection, since, while related (and mutually dependent in a way that would have been abhorrent to the Viennese), they are still distinguishable kinds of problems. It seems to me that in the case of meta-ethical discussion, we ex-positivists have a special tendency to move too readily from one to the other, and furthermore—a mistake which we would be unlikely to make in discussing chemistry or pure mathematics—to conflate a third component, namely the component of impure pragmatics, so that we try to handle all three tough problems simultaneously without separating the difficulties that characterize them. Is it not true, for example, that someone with your and my positivist humours in the blood, when asked to reflect upon the (possible) truth-status or correctness—I do not say “validity” because we can take care of that by reference to derivability within the ethical formalism when conjoined with the biological and social facts—of a primitive axiological statement, moves reflexly and immediately to the surely distinguishable question in pragmatics “But what if Hitler

doesn't agree with me about cruelty being wrong?", and hence quickly into discussion about anthropology and ethical relativity, and persisting unresolved disagreement between Kantians and utilitarians, and so forth. It must be obvious to you from all that has preceded in this memo *supra* that I have no intention to minimize or side-step the seriousness of this question, nor do I mean to avoid the possibility that the origin of this rock-bottom unresolvability of primary ethical "Boo-Hoorah" commitments may ultimately lie precisely where the Viennese and others before them thought it did, that is, in the metaphysical lack of any objective "extra-mental" *state of affairs* to which the ethical judgment may "correspond" or "fail to correspond," with respect to which it can be properly said, in the strict use of language, to be "true" or "false," and which external corresponding state of affairs would be capable of generating or producing or causing or giving rise to evidentiary experiences that would be persuasive to someone who disagrees with us initially. I don't want to minimize or avoid or obscure any of that, which I take in deadly seriousness. It's just that we positivists move so quickly and automatically and, to be blunt, unthinkingly from the Big Stomach ache about "nothing being out there for the ethical sentence to correspond to," to the demand for an account in the semantic metalanguage of the meaning and truth of ethical predications in terms of nonethical descriptors (see criticism of this in Section 30 *supra*), then to the epistemological problem of evidence (especially Hume's Guillotine), then to the difficulties of influencing a desperately wicked or morally blind individual—all of which steps we are likely to slide through in 30 seconds or less. And I maintain we would not do that if we were considering the ontology of statements in Freudian theory or molecular biology or atomic physics, would we?

But we were asking what it is that the ethical objectivist wants so badly to be able to say that isn't included in the rather impressive list of "cognitive, indicative, arguable" things under Subsection (b) *supra*, all six of which categories leave plenty of room for assertion, denial, rational argument, formal and empirical reasoning, and the like. Now this question about what the ethical objectivist wants or needs to say that he can't say on a pure postulate view (pragmatically, on a pure personal commitment view) is a question about his statements in the metalanguage. Obviously he can say whatever he wants in the ethical object language, and for the theorems he can also argue for them, i.e., can show that they flow from the pure postulates he has "adopted." And we have seen that while in one sense the primary axiologicals are pure postulates, in another sense they are also corrigible, because he makes a resolution to adopt such postulates as will generate theorems that, when applied to fact situations, will tend statistically to accord "reasonably well" with the Prichardian ethical intuitions *in concreto* of his bootstrapped ERDERVE "ethical observers," to really push the analogy to empirical science! The postulates are pure postulates, but that doesn't mean that they have not been chosen with an eye to their consequences, just as Russell and Whitehead point out that they have chosen their postulates so as to be able to generate what is already believed about logic and mathematics. And there is nothing about a metalinguistic statement that a certain kind of sentence (in

this case the distinctively and uniquely ethical predicate ‘ought’) is a primitive that requires a denial in the metalanguage that the sentence is true or correct. That is, if I say in the ethical object language that “one ought not to be cruel,” and this is one of the postulates rather than one of the (derived) theorems of my ethical system, there is surely nothing about the general principles of logic and epistemology that requires me—if I inform you that the epistemic status of this statement about cruelty is that of a pure postulate and that its status in impure pragmatics is that I am committed to it—to say in the semantic metalanguage, “This statement is false,” or even “This statement is neither true nor false,” or “He who believes the contrary statement ‘One ought to be cruel,’ or the contradictory statement ‘One may be cruel,’ is ethically correct, or at least is as ethically correct as I am.” What I am emphasizing here is that the question of how I would go about persuading him (a question in impure pragmatics) or what I would say about the situation if I find some people who, despite ERDERVE, remain unpersuadable (a complicated question in impure pragmatics and epistemology) and how I give a causal account of the occurrence of such persisting disagreement among ERDERVE ethical reasoners (a Copernican Turn question about ethical knowledge)—all of these are deep and fascinating questions, and they may bear in some complicated way upon the pure semantic question of truth value, I don’t know about that; my point here is that none of them is *the same question* as the question whether an ethical objectivist is somehow estopped, on a pure postulate theory of primary axiologicals, from asserting in his ethical metalanguage, “This primary ethical postulate is true.” It is not clear to me whether you can stop somebody from saying that a postulate is true or correct or sound or whatever we want to say prior to his providing a thoroughly satisfactory elucidation of the notion of truth in general and the notion of truth in the particular domain. It is my impression that if we did that, we could also prevent people from saying anything in logic or in chemistry or in psychoanalysis at the present time, could we not? But in any case, aside from what sort of affirmative reconstruction of the kind of ethical correspondence theory of truth the objectivist may be able to provide us with (or refuse to provide us with as a condition for asserting what he wishes to assert), he cannot be compelled to assert in the ethical metalanguage, “Since my primary postulates are pure postulates, and since I have adopted them by an act of personal commitment following ERDERVE and not by syllogistic inference from nonethical propositions formulated in psychologese, *therefore* your primary postulates, which are inconsistent with mine, are also or equally correct or true.” A kind of nonsensical remark, anyway, isn’t it? What it would mean is what the positivists said, that *none* of the postulates can be spoken of as being correct or true. I just want to avoid mixing up the matter of persuadability (and the explanation of the fact of non-persuadability) with some kind of forced metalinguistic ethical relativism in the way that positivism tended to argue with entirely too much ease and, as I now think, dogmatism.

While I don’t want to claim that ethics is like logic, for the same reasons that you don’t, in some respects this particular difficulty is, in ethics, somewhat like the difficulty

in logic, as your paper "*De principiis ...*" suggested, although as I recall with some ambivalence on your part. You and I many times discussed what you do when a mystic says—as some of my daughter's hippy or Zen types will say in so many words—that he "doesn't believe in logic" or "cannot be made to reason about such and such." We know as a matter of psychological fact that he can't be "made to" reason in any way, short of putting him on the rack and forcing him to say the words whether he believes them or not. Suppose I explain to somebody the Lewis-Langford hook notation, which is formally introduced by them via their primitive diamond notation representing *possibility*, but of course (as is always true in such treatments) is surrounded by a bunch of common sense and intuitive discussion in the embedding text, so that the reader can know what he is "really talking about." Let's not hassle pure logic in this connection. I write down, for instance, a tautology in the propositional calculus which I show by truth tables to be a tautology. Then I am permitted to substitute the hook for the main horseshoe in that tautology, right? Now what I say in the metalanguage is that this is a tautology (if I am talking in terms of truth functions) or that it is a strict implication (if I am talking modal logic). But in either case I also may say that a mystic or a Zen Buddhist or a schizophrenic who refuses to obey it or who claims at least not to understand it, has "got something wrong with his thinking." That is, we say of him that he "thinks illogically." But of course I need some pure postulates to do all of this, and while there is a choice open to me as to which statement schemata I pick as postulates and which come out as theorems, there is no way to avoid having some pure postulates. I find there are 5 pure postulates in Alonzo Church for the lower functional calculus, and 9 are required for the second order functional calculus. It wouldn't surprise me if one could squeak by with not much more than that for a "satisfactory ethics." Nobody (except maybe Carnap in his super conventionalist mood, but not later) wants to say that a logician who accepts these as pure postulates (although with an eye to the theorems he wants to derive) is committed to saying in the metalanguage that they aren't true or correct or valid. Some people, I suppose in fact the great majority today, would on the contrary want to say in the metalanguage that these logical postulates are true or correct or sound or valid, and they would mean more than the formal meaning of that last word in the sense that all of the theorems of the deductive system are valid and the theorems, if I recall the way you guys talk, include as a sort of degenerate case, the postulates themselves. (I guess I have been careless about this in some of the text in sections *supra*, where I have usually if not always meant by the word 'theorem' a derived theorem, i.e., *not* one of the postulates. But no harm done, I dare say you understood me throughout to mean this.) We could get into quite a discussion about which of the postulates of a logic one could change and still have a logic, a topic that I know has interested you for many years and led you not to like to speak of three valued or many valued logic as quite the same sorts of thing as the ordinary two valued logic. Without expertise, my sentiments here are with you, because we can't even apply one of the so-called many valued logics without using our normal two valued logic when we look at the truth tables (or the derivation chain, or whatever

decision procedure is involved in testing whether a particular formula is a theorem or not, and so on). Surely nobody has ever proposed a many valued logic in which such a question as, for instance, “Is the formula F well formed?” is neither true nor false but something in between or in limbo? There are some deeper connections about language (or, as I would prefer to say these days, about thought) which do not seem to be “arbitrary” or “capricious” or “volitional decisions” or, God save the mark, “conventions.” I don’t particularly like the old positivist move that we used to make when confronted with the fictional person who won’t be convinced by a clean one step *modus ponens* or won’t permit us to rely upon the Rule of Detachment, which was to say, “Well, of course, he himself can’t coherently assert anything then.” Or, even worse, “He himself will only utter meaningless statements.” Because of course he could turn his willingness to follow a logical principle on and off as a function of caprice or sun spots or whose ox is being gored or whose axe is being ground, which is what most slopheads tend to do in ordinary life anyway! And I can’t for the life of me imagine how you would go about “refuting” (worse still, compelling assent, i.e., influencing propositional and attitudinal behavior) in a fanatical flat-earther or hard shell Baptist or whatever who in effect said, when pressed sufficiently, that he thinks in accordance with those axioms when talking about poker chips but not when he is talking about the Trinity, or our knowledge of the external world, or the Cosmic All. It won’t do to tell him that he is being inconsistent in his refusal to apply the rules of logic consistently, because by his initial step he has shown us (what he may or may not be able or willing to say to us in the metalanguage under criticism) that he does not count consistency as one of the necessities or, in some areas, even as one of the virtues. Witness Martin Luther using scholastic logic to refute certain Roman doctrines, and yet also referring to “die Heure Vernunft” or taking those scriptural texts which on the one hand clearly assert predestination along with other texts, such as the passage in Peter that “God will have all men to be saved and come to the knowledge of the truth,” thereby holding to single predestinationism as against Calvin’s (consistent but dreadful) double predestinationism. Luther was not feeble minded or insane, and he was well educated in medieval logic as well as in the almost positivistic thinking of d’Ailly and Ockham—yet he turned the application of concepts and syllogisms on and off in the service of his polemics. Needless to say I do not view any of this as admirable, I merely point out that we have here another example of rock-bottom collision which cannot be resolved, nor can it be explained by attributing madness or stupidity or ignorance to the opponent. It’s simply that he just won’t have it our way, and that’s all there is to it. I am afraid the most we can say to somebody who doesn’t feel the compulsion of the hook is that he doesn’t think logically and that he won’t get to the right results if he persists in thinking illogically. The ethical analog is, I think, quite obvious: What we say to somebody who doesn’t apply the same rules of conduct as a function of whether he or the other fellow is in the agent versus patient relation of an ethical predicate is that he doesn’t “reason ethically” and, therefore, if he persists in these bad habits and acts in accordance with them, he will not “behave morally.” I repeat that I am not saying there is

no difference between logic and ethics, but I am saying that *on this score* (not, probably, on others) we are just as entitled to say in the metalanguage that someone who disagrees with our own primary ethical postulates is ethically mistaken and if he behaves in accordance with his wrong ethical beliefs, morally derelict, as we are in saying in logic that a person who doesn't accept the postulates of inference that cut across different calculi and even so called "different logics" is thinking illogically, is making a mistake about the relations of propositions (I would now prefer to say, the relations of ideas); and that, if he persists in these bad habits of thinking, he will come to mistaken conclusions about the way things really are. These remarks are, I realize, largely repetitious of the point I made in Section 23 *supra*, especially the remarks about logical principles ("First transition premise," pp. 22–25), and I fear I am not adding much to what was said there except maybe a slight difference in flavor and emphasis.

Since I alluded to the horseshoe and the hook a couple of paragraphs above, this is a good a place to say something about that as any (I am dictating this last section after almost all of the preceding material has been typed in final form and proofed). It has to do with the pragmatic dimension involved when we speak of some kind of discourse or symbol as being 'prescriptive.' There is the strong morally laden meaning of this, but (as I think I mentioned above in the discussion of "What is a rule?") everyone recognizes a prescriptive element in anything called a 'rule' whether or not legal coercion or moral condemnation would or should follow from an infraction of it. I said that we would not consider somebody who fudges or cheats at Solitaire either criminal or morally wicked, although cheating at Solitaire seems to most of us kind of pointless. If a logician or mathematician fiddles with a new calculus with no intention to apply it to human affairs or even to publish it, we would still be able to make the distinction between his following or violating one of his own calculus postulated rules, not to mention violating one of the rules of what Carnap called the *general calculus*—that is, rock-bottom pervasive logical rules which are not peculiar to different calculi but cut across them and are, on Carnap's logicist position, consequences of the basic postulates of general logic itself. If I put down a sufficiently complicated expression in the propositional calculus so that only a logistic genius could intuit in a flash whether it is or is not a tautology, the main horseshoe will, to most of us, carry a certain impulsion or flavor of "push" from left to right, but after enough immersion in discussions of the paradoxes of material implication, I think even an undergraduate philosophy student would not feel this as strongly as one feels it when he first hears that the horseshoe stands for the common language "If ... then ..." and prior to having it explained to him that it just means that it is not in fact the case that the antecedent is true but the consequence false, and hearing such homey examples (invented to soften up his ordinary language resistances) as "If McGovern wins, I will eat my hat." If I write down

$$(p \supset q) \cdot p : \supset : q$$

most of us will feel a very strong push to the right, probably as strong as if I wrote

$$[(p \supset q) \cdot p] \prec q .$$

We might feel a little stronger pull reading the second form, but not much. Why is that? It's because the first form, in material implication, is immediately recognized by us as a tautology, and while the hook is "stronger," we know that any time an expression in material implication is tautologous in the truth tables, we can substitute the hook of strict implication for the main horseshoe in the material form, right?

But now if I write

$$[p \supset (q \cdot r)] [(r \vee s) \supset t] \supset [p \supset t]$$

I think for most of us that has a considerably weakened "force" moving us from the left to the right side of the main horseshoe, because it isn't intuitively obvious in "If... then..." ordinary language. Hence it isn't obvious without running through the truth tables that it is a tautology; whereas if I replaced that big horseshoe by a hook, we would still feel a strong impulsion to move from the left to the right side, although of course it might occur to us that the author was an incompetent logician and had made a mistake.

You probably think all this hasn't anything to do with much of anything, let alone the metaphysical status of ethical truths, but I think it helps a bit—by reflecting on a non-ethical example—to keep separated the component of pragmatics that I mentioned toward the beginning of this section as too often conflated with the "correspondence question" put in the semantic metalanguage. That is, we don't recognize any logical difference between the horseshoe as it appears in the simple example above and the complicated one, but we are not distressed to notice a clearly introspectable pragmatic difference in the impulsion in the two cases. We can explain the difference psychologically in terms of "obviousness," an explanation which wouldn't work for a sufficiently acute and practiced logician who would also see the more complicated one as tautologous and hence immediately feel sort of "hook-like" about the main horseshoe, the way we all do about the main horseshoe in the simple one that merely instantiates the propositional schema of *modus ponens*.

This pragmatic dimension corresponds to analogous but distinct pragmatic dimensions in using our observation language, as when we feel impelled (or if under oath or in the scientific laboratory one could almost say "compelled") to pronounce the semantically appropriate color word or number expression in the presence of a bluish flame or a dial needle lying between dial numbers 6 and 7, we experience an impulsion to perform an action which concludes an ethical ratiocination by terminating in "I ought to report this to the FBI." I repeat in this connection that it is not, in my view, an illicit reliance upon the usages of vulgar speech but one of those perfectly kosher instances of "following Webster" to say that it makes no sense to ask the question "Why ought I to act obediently to an ethical rule?", since this collapses to the silly question "Why ought I to do what I ought to do?", the only answer to which is the moral tautology "One ought to do what he ought to do." Of course to show that it is somehow *to my advantage* to do what I

ought to do is a completely separate question (although Plato seems to have been mixed up about this, as Prichard points out in his famous 1912 article), and for my part it seems blindingly obvious that it is *not* always to one's advantage to do that which he ought to do. That is, it is sometimes to one's advantage to behave unethically. Questions about the compulsive influence of a sincerely believed sentence containing the distinctively ethical predicate 'ought' are questions of conditioning and learning and motivation, and answers to them are, in my opinion, exactly the same as answers to questions such as how it can come to pass that a person tends to utter a token of the predicate 'Red' in obedience to a semantic rule, "'Red' means *red*," when he is having a visual experience of red quality. I can't make any more sense out of a metaphysician's efforts to generate an impulsion to act in conformity to a moral rule as if it were a philosophical problem of derivation (which I maintain it is not) than an epistemologist striving to generate from pure logic, or even from pure logic supplemented by the semantic rules of a color language, the psychophysiological disposition to name a color or to have any of the other associated experiences and impulses.

There is of course a striking difference between the irrational fellow who won't start with the postulates of the propositional or functional calculus and the sociopath who won't start with ethical postulates forbidding cruelty or promise breaking, and I have no wish to fuzz up this difference or to paint it over. The difference lies in the fact that, while we can't compel the illogical person from having recourse to those very rules as it suits his fancy (because we can't get mileage out of convicting him of an inconsistent or capricious application, inasmuch as he has apparently repudiated consistency), we are able to say to him *that* he is contradicting himself, that he can prove all propositions including those he doesn't like, that he himself relies upon the Rule of Detachment. All of these things can be said, and we can agree among ourselves that they convict him of a gross incoherency even though he may succeed in maintaining a blissful unconcern about it. But of course we cannot do any such thing when it comes to the rock-bottom postulates of ethics, unless we can hang around and catch the sociopath making ethical judgments, including second order ethical judgments about somebody being "unfair" because he turns his own ethics on and off depending on whose ox is being gored. I think we *could* usually do just this, and that's one reason I think that even a nut like Hitler is not precisely in the position *vis-à-vis* an ethical critic that classical Vienna positivism tended to put such people in. That is, I think Hitler's moral reasoning is shoddy and incoherent, much of it not only because of the ethical postulates but because of crazy empirical ideas about Aryan blood and the like. Let us imagine a person who, unlike Hitler, could think straight and didn't have his head full of crazy garbage about empirical matters, and who still collides with us on a rock-bottom ethical proposition. Suppose it turns out that by going through all of the cognitive kinds of activities mentioned in Subsection (b) immediately *supra*, we don't detect him in phony derivations or inconsistencies or in insincere reports on his Prichardian intuitions with which he claims his ethical postulates are consistent and capable of deriving—then we seem to be stuck with saying his primary ethical

postulates are different from ours, which is a weak statement compared to what we can say in logic about somebody who refuses to go along with the hook or to prevent us reliance on the Rule of Detachment. So I still think, with the old Vienna boys, that there is some basic difference here in reality status, but it certainly seems hard to say what it is.

There is one ontological stomach ache which I intuitively believe is somehow different between ethics and logic (as I believe it is very different between logic and aesthetics, where I take the old-line positivist position, probably because I am myself not very artsy crafty!) but which I never seem to be able to put into words and therefore am suspicious about, wondering whether it is also merely a hangover of Victorian materialism and Vienna positivism. One keeps thinking that somehow there has to be a “something” (an entity? a physical something?) out in the world external to human minds and outside of human society, in order for an ethical sentence to have something “out there” to which it can “objectively” correspond, otherwise an ethical assertion would not be “genuinely indicative.” I must confess I can’t get too excited about whether we say an ethical sentence is indicative, although I would be inclined to resist any analysis that said it was wholly imperative, for the familiar reasons, including those which have been advanced by jurists against Austin’s theory of law as being commands of the sovereign. Since I take the notion of *obligation* as a primitive, I find myself unable to get excited by analyses purporting to reduce it to something, whether to imperatives or expletives or anticipatory guilt feelings or whatever else. There is a whole cluster of dispositions and antecedents to a state of mind a morally aware person is in when he cognizes that he ought to do something or refrain from doing something; and I see no more reason for trying to unpack the ethical primitive ‘ought’ (which in the embedding text partially interprets the meaning of the obligation vector discussed in Section 29 *supra*) in terms of the child’s development of the moral sense by a mixture of models, precepts, and differential rewards and punishments (the three main sources) or the outcomes such as anticipatory guilt or shame, actual guilt or shame, the impulse to ask forgiveness, to make restitution, to punish oneself or submit to punishment as an expiation, to make New Year’s resolutions to strengthen one’s will or shape up one’s character so as not to do it in the future, and the like, than I do to try to explicate a color predicate or a more complicated expression in the ordinary physical thing language by unraveling the whole history of a person’s verbal conditioning and the current state of the reinforcement schedule provided by his “verbal community” as Skinner calls it. This is an interesting thing for the psychologist to do, and I don’t dispute that it is something at the interface between epistemology and psychology, but I would strongly resist the idea that the task of the philosopher of perception and, even worse, the task of somebody constructing a color vocabulary or a proposed set of rules for conversation among paint chemists or ophthalmologists or architects, should feel compelled to pack all of that stuff into his conventions. I think this is simply a mistake, and for some reason it is one which we are much more prone to commit in ethics than we are in other fields. I suspect that the explanation of this also lies in justificationism and positivist reductionism to nonethical

descriptors and therefore ultimately with the constant threat of Hume's Guillotine, in the sense that due to Hume's Guillotine we think we must be talking nonsense if we say something that is not reducible to descriptors, and the only descriptors that seem plausible candidates when we are talking ethics are human beings' motives and feelings about actions; hence the temptation to define, reduce, or otherwise explicate distinctively ethical notions by motivational-emotional notions.

But I digress, although usefully. I was saying that, despite the important differences between ethics and logic, I see one respect in which they are similar and it may be identical as to the ontological problem of "out there-ness." We would not think of asking Lewis and Langford, "Where, out there in the world outside your and my mind, is the entity designated by the hook of strict implication?" Yet, leaving out our schizophrenic or our insincere logic chopper for the moment, confining ourselves to people who understand the notation, we all want to say that if, given the fact that a certain formula (say, *modus ponens* as above) is a tautology in material implication, the main horseshoe can be replaced by the hook of strict implication, then a person who asserts the antecedent and denies the consequent is just plain wrong, is making a mistake, is about as far out of kilter with the way things are as you could possibly get, even more out of kilter than somebody who believes that the earth is flat (which, after all, while wildly implausible is nevertheless an empirical sentence and consequently, we usually think, not of such certitude as a sufficiently simple theorem in formal logic). That is, we want to say that his thinking isn't in accord with reality, although the reality involved here is the reality of logical relations rather than the reality of stones and butterflies. Remember, we are not discussing the pragmatics of persuasion; and we are not discussing an *affirmative reconstruction* of the ontological status of an ethical predicate (i.e., how the "correspondence theory of truth" applies here); we are discussing a restricted question about, shall I say, the matter of "objective reference," in which one finds oneself hard put to visualize (goodness!) or imagine or conceive a "something" which is "out there in the external world" to which the predicate "corresponds." I realize I am writing as if I had a very simplistic correspondence theory in mind. I am doing that on purpose, because it seems to me that there is a kind of simplistic correspondence theory *demand* in the back of the positivist's mind when he insists that the ethical objectivist explain to him in the ethical metalanguage (and, to add insult to injury, by the use of nonethical descriptors only, see Section 30 *supra*) what kind of outside entity that is. We have repeated over and over again that it is not an entity in the sense of a structure or event or disposition. It is a relational entity between moral agents and other moral agents or other sentient beings that are the proper objects of ethical obligation, period. Now, while logic is different from ethics and you can get into all kinds of trouble by being illogical, troubles more pervasive, fundamental and looney than the troubles you get into by being unethical, I am here addressing myself only to that narrow aspect of the correspondence stomach ache that we suffer as crypto-materialist positivists brought up in Western scientism as our intellectual tradition. If somebody says that he won't permit us to employ a notation to

designate a primitive relation like obligation unless we can explain to him in our metalanguage what kind of an entity it is that is “out there”—sort of like G. E. Moore imagined his opponents to be insisting upon a reduction of his primitive ‘good’ to some kind of a “naturalistic predicate”—it seems to me we are entitled to ask him what meta-meta-principle he is here invoking against us, and threatening to use to decapitate us with his variant of Hume’s Guillotine if we don’t deliver the goods? And it seems to me that to make his objection stick *in this narrow special way*, he would have to come up with some kind of statement that would say, roughly, “A notation for a relationship is forbidden unless it stands for an external entity that could in principle be pointed at or perceived or located in space time,” which is, it seems to me, a really crude form of the ‘Fido’-Fido theory of how expressions designate. The minute he does that (which I think we can force him to do by telling him that we aren’t obliged to deliver the goods by way of answering a meta-meta-question about ethical predicate designations unless he can tell us what it is that he wants and why he is entitled to demand it), we clunk him over the head with his own question as applied to things like the Lewis-Langford hook. Surely there isn’t anything “out there in space time” that “corresponds to” the hook of deducibility, is there? Please don’t tell me that it’s not out there but it’s still for real and ethics isn’t. I’m not making that kind of a point here. I am rebutting a criticism, not offering an affirmative reconstruction of ethical ontology. If I were an old fashioned type of ethical realist or objectivist, somebody like Urban, I would here start talking about the Realm of Values, that is, a domain of entities and relations which has its own kind of meaning and truth and its own criteria for validity and knowledge, and analogize it to the realm of facts and to the realm of mind and to the realm of formal relations in pure logic and mathematics, each of which has its distinctive kind of truth and reality although they all share the concept of validity insofar as they are rational at all. I am not going to permit myself that kind of a sermon, but I understand how tempting it is, because if a positivist ethical relativist emotivist critic keeps pushing on one’s inability to, so to speak, “point to” a lump of sludge out there in the world located 12 feet southwest of the tree stump as the entity designated by the distinctively ethical predicate ‘ought,’ I can see why it gets pretty tiresome and one begins to have an urge to resort to *tu quoque* by asking about the realm of logical and mathematical truth how they would fare under this same kind of attack.

I repeat, I am not here trying to make an affirmative reconstruction. I think it likely that the ethical absolutist wants to be able to say something that he cannot say about the objective truth status of the postulates, something over and above their status as pure postulates and, in pragmatics, his adoption of them as personal commitments. But I merely want to insist that one who utilizes concepts such as deducibility or who engages in discussion about the merits of Brouwer in the matter of the Excluded Middle as applied to nonconstructive proofs in transfinite arithmetic, given the state of that enterprise, is not really entitled to the simplistic kind of correspondence principle as used against ethics.

What then is left that we want to assert in the metalanguage? I leave it to you here, as I have trouble saying just what is “missing” that the objectivist wants to say—although I cheerfully admit there is something lacking.

SUMMARY THESES

1. What is the definition of the term ‘ought’?

My answer: It should be taken as a primitive, probably the sole ethical primitive. If ‘good’ is an ethical primitive (which I doubt, suspecting it to be definable in Utopian psychologese, despite G. E. Moore), then there will be two ethical primitives, neither definable in terms of the other.

2. What is the status of propositions containing the distinctively ethical predicate ‘ought’?

My answer: Some are postulates, others are theorems. Presumably which are which is to some degree interchangeable in ethics, as is the case in logic, mathematics, economics, and so on. But in addition to theoretical convenience, it may be desirable to give postulational status to those that, upon reflection, are most obvious to the mind.

3. What is the relation of ethical postulates to non-ethical “descriptive” statements?

My answer: As a matter of formal logic we cannot derive either from the others if we adopt a justificationist program. We accept Hume’s Guillotine, but it doesn’t bother us. If, however, we are not justificationists, then we abandon the wish to have a linear syllogistic inference chain starting from ‘is-statements and ending with an ‘ought’-statement. Instead we examine whether the *set* of ‘is’-statements conjoined with the *set* of accepted ‘ought’-statements generates contradictions, contains gaps where (realizable) moral cases are not decidable, violates our intuitions when applied to concrete cases, and the like. If one abandons a justificationist epistemology for ethics (as one has already abandoned it for science) and therefore no longer requires some kind of linear deductive inference from “is” to “ought,” but rather postulates (conjectures, adopts—existentialist moralists would presumably say *commits oneself to*) a particular set of ethical postulates, then on a non-justificationist view this set of ethical conjectures is evaluated meta-ethically in a complicated way. Despite Hume’s Guillotine, it does relate to “facts,” such as the basic facts of human nature, the current situation of the society, the direction of its change, and the like. No new theory of truth is involved here except, of course, the primitive ontology: An ethical obligation is an undefined notion regarding the distinctively ethical relation of the parties. We have already rejected the idea that this has to be “defined” or “reduced to” some kind of nonethical, naturalistic, physicalistic, ordinary descriptive predicate. The evaluation of the conjectured set of primary ethical postulates takes place by a complicated combination of the coherence and correspondence theories of truth, as is

also the case of scientific theories. Although here again it is to be emphasized that complete reducibility to scientific predicates in the usual sense is not only not required but is repudiated. That is, we *do* accept Hume's Guillotine (as re-interpreted in our complicated system) and the uniqueness of the primitive ethical predicate "ought." But we can ask about the conjectured set of primary ethical statements whether they cohere, whether they cover the ground reasonably well, whether they lead to consistent moral conclusions about real cases that arise or that can plausibly be imagined to arise. In terms of the correspondence theory of truth we ask whether the consequences for real cases square with our reflective Prichardian intuitions *in concreto*, and also whether the postulates themselves seem to agree with the abstract Prichardian intuitions of ethically sensitive, reflective, intelligent, rational, fair-minded and generally informed individuals.

4. How can it be shown that somebody ought to obey the moral law?

My answer: This is the only one where I will rely upon the ordinary language argument that it's what we mean by being moral, i.e., "I ought to do that which I ought to do" is a tautology. (This is a *genuine* tautology, as its form shows. It is *not* a fake tautology, such as the ordinary language solution of Hume's problem saying "That's just what we *mean* by 'being reasonable' in empirical contexts.")

5. How can it be shown that it is the part of wisdom, or in some sense advantageous, for me to obey the moral law in all circumstances?

My answer: It is fruitless to bother trying to prove this because it is false. Absent special theological propositions (not including classical Protestantism), it is *not* always to my advantage to obey the moral law. Sometimes it is distinctly to my disadvantage. Sometimes it might even be fatal (e.g., to refuse to fight in an unjust war, as a result of which the dictator has me shot).

6. How can it be explained that people are impelled to obey the moral law, especially when it is to their disadvantage?

My answer: As Aristotle says, this is a matter of habit and training; or as Kant says, it is a matter of "respect" [for the moral law]. Whether "practical reason" can itself constitute an inducement, I leave open, although my views on cognitive psychology would be that at times we are impelled because it is morally rational similarly to the way we are impelled to obey the rules of the syllogism in a nonethical domain.

7. What is it for a moral judgment to be incorrect?

My answer: A moral judgment can be wrong (unsound, incorrect, invalid, counter-ethical) in several ways. These ways are distinguishable, but not totally unrelated. Their relations are complex and remain to be clarified, but these problems are not unique to ethics, although probably more severe for ethics than for, say, chemistry or genetics. The syntactic, semantic, pragmatic, ontological and epistemological issues must be distinguished, but not treated as if wholly unrelated, nor as if one of them

(e.g., epistemology) were logically prior to another (e.g., ontology). A more satisfactory account of the concept *ethical mistake* will hinge upon development of a satisfactory metatheory of truth generally, especially as regards the (likely, hoped-for) fusion of “coherence” and “correspondence” theories and their integration with the pragmatics of human decision-making. Pending such advances in general philosophy, we can only list the several *different* ways in which an ethical judgment can err. It can err in some of these and not in others, as is true for judgments in non-ethical domains:

- a. As a pure postulate, it can be morally incorrect, FULL STOP (as Alburey Castell used to say). If a satisfactory ontology of ethics can be developed, one supposes that this would be *the* meaning of “erroneous ethical judgment,” the privileged status meaning that the others would depend upon—if we accept physics as our paradigm cases of objective knowledge. But one should proceed cautiously at this point.
- b. As a purported commitment, it can be psychologically inaccurate (I lie, I’m a hypocrite, I’m a poor introspector, or a careless self-behavior-observer.)
- c. While accurate for me now, it would change with ERDERVE.
- d. As a postulate, it is mainly responsible for generating theorems that offend the ethical intuitions *in concreto* of ERDERVE-qualified persons.
- e. As a postulate, it is inconsistent with the other postulates (collectively) I hold.
- f. Although not inconsistent with the other postulates I hold, when it + they + facts (of mind and society) are conjoined, contradictory moral theorems result.
- g. Apart from theorems, as a postulate it is counter-intuitive to ERDERVE-persons and will become so to me given adequate ERDERVE.
- h. As a purported theorem, its derivation from the postulates is fallacious.
- i. If it is an “ethical protocol” (intuition *in concreto*) I react differently depending on which position in the two-place obligation function I am occupying (agent vs. patient).
- j. If it is a conclusion derived from ethical theorems conjoined with particular (situational) facts and means-end relations, the “facts” relied on are incorrect, i.e., they are not facts.
- k. If it is a conclusion derived as in (j) preceding, the means-end relation relied on is empirically false.
- l. In applying the net obligation vector to a moral choice, examination of my other decisions shows the present selection to be arbitrary, capricious, unprincipled, way out of line with the total set.

- m. A *prima facie* obligation stated qualitatively is derived satisfactorily from ethical theorems + particulars + empirical means-end correlations, but its collision with another is ignored (Fallacy of Neglected Aspect, as old Castell's text calls it.)
 - n. A fact-situation is not "well-subsumed," a common issue in law. The formalization of ethics in Section 29 *supra* was a crude start, but even if carried through would, like all mathematicizations of biological or social science, be an idealization. In real life ethical disputes, the characterization of responses in the first matrix and of morally relevant response consequences in the second are "unrealistic" formalizations given the real properties of mind and society. Hence questions such as, "Is a motorized child's racer a 'motor vehicle' in the sense the ordinance forbids from the park grounds?", with which lawyers are familiar, can be expected to arise in a formalized ethics. *A fortiori* they arise—and result in persisting disagreement—in our contemporary ordinary language ethics. Exactly what kinds of subsumptions are erroneous—what it *means* for an "open concept" to be misapplied—is unclear. Yet *that* one can mal-subsume seems evident, as if I call a bearded hippie "clean-shaven."
- 8. What can an ethical dispute be about when not about empirical means-end connections?**

My answer: To this standard positivist question I make three responses.

- a. The question, while legitimate and weighty, is (in my experience anyway) typically put with the hidden agenda to suggest that no "really *ethical*" disputes are possible, were it shown that (as I of course deny) only "means-end" empirical issues remain after we see that all the rest is sheer commitments and these are unarguable. It seems to me that such a view would, if applied to other domains, have a funny trivializing effect on them as well. We would not, I think, dismiss a discussion in physics as not really being "physics" if it turned out that it ultimately hinged upon, say, the accuracy of a certain laboratory instrument, or whether it is possible to provide a non-arbitrary mathematical basis for excluding "wild solutions" (divergent integrals) in quantum mechanics. Such issues are not neatly pigeonholed as being "really mathematics" or "really physics," and no one feels an urge to do that. In the social sciences, we would not say that a debate about the genetics of schizophrenia wasn't "really about genetics" if it hinged on the validity of a Rorschach sign, or the therapeutic specificity of Prolixin in cases of pseudoneurotic schizophrenia. Assuming we could work out the matrix formalism adumbrated in Section 29 *supra*, an ethical dispute over a particular fact-situation would have to involve the complex network of relations among the several matrices, and the empirical consequence matrix would be examined in its rôle in the system. Real life ethical disagreements often turn mainly—sometimes wholly—upon empirical means-end contingencies concerning which the disputants disagree. But the conduct of such conversations is guided by their mutual

understanding that the whole network provides the background—the “motivation” as mathematicians sometimes say—for what is said and why we count it as “worth saying.” I would therefore insist that *even if* one conceded that nothing but the empirical means-end probability matrix is in real dispute, the usual emotivist implication that this would declassify the argument from being “about ethics” is misleading and positivistically tendentious.

- b. But of course the preceding subsection (Summary 7) about ways in which an ethical judgment can be erroneous, and the main Section 37 *supra* about what assertions the would-be objectivist can make on a “pure postulate-personal commitment” view, provide the chief reply to the question.
- c. While ERDERVE-criteria are mentioned in 7 preceding and hence in a way covered by answer (b), there’s an important difference between *mentioning* ERDERVE and *providing* it. I have an impression that some ethical discussions include “appeals to the moral sense,” when we invoke a cluster of notions that include and connect morally relevant act-consequences, moral labellings, and the ethical relation ‘ought’ itself, the whole schmeer being presented in a hortatory fashion. The reflex positivist criticism of such a proceeding is to condemn it as “mere emotive appeal,” which is an odd complaint from one who thinks that primary axiologicals are personal commitments anyway, don’t you agree? Because if the Boo-Hurrah theory is partly correct (or even wholly correct as to the pure postulates), it would seem that appeals aimed to shift my “primary ethical feelings” from Boo to Hurrah (or conversely) would be the appropriate strategy. So I suggest that discussions of this kind are both explainable and reasonable—they aim to strengthen some of the listener’s commitments by *describing and evaluating* actions and their consequences, in the context of a larger network that is, as usual, a mix of moral and psychosocial beliefs. “Think how *you* would feel if your boss...” is not a forbidden move in the ethical language game.

9. What do we do with Hitler?

My answer: First, we don’t vote for him; secondly, after he has taken power and breaks the political rules, we kill him. Of course we try other, less bloodthirsty, methods of removing or inactivating him first, as we got rid of Tricky Dick Nixon without violence. But Hitler would not have resigned, and the Weimar Republic’s spineless judges would not have decided as our Supreme Court did about the tapes. So when Hitler persists in murdering people and threatens to do us in for interfering, we have to kill him. These (perfectly sound) answers are not satisfying to the positivist, who brings up Crazy Adolf as a horrendous example of how subjective ethical judgments are. For my part, I am not in the least impressed with the sad social fact that a nut takes power. The positivist’s reference to Hitler is merely a “punchy” way to emphasize what *he*, the positivist, concludes from our admission that ethics is

pure postulate, namely, that “therefore” *one moral postulate is as good as another*. Hitler’s “taste” permits, says the positivist, liquidating six million people, and who are we to fault someone for acting according to his tastes?

In reality, what we would try to “do with” a Hitler includes *all lines of inquiry* (facts, correlations, ERDERVE outcomes, coherence, derivability, subsumability) that are possible for any of the cognitive issues listed in 7 *supra*. Reflecting on the course and outcome of such a process, we see that the positivist’s example turns out to be poorly chosen for his purposes (i.e., to show “nothing can be said or done”), because any such efforts with the Hitler example will quickly reveal his gross incompetence to engage in such a discussion. The positivist wishes, by his Hitler example, to highlight the idea that when confronted with a person who adopts ethical postulates differing from my own, I “can’t do anything,” because, says the positivist, the postulates are not arguable, empirically testable, and so on. The picture intended is that of a sane, informed, rational, open minded person who (mysteriously, arbitrarily, capriciously, but *quite legitimately*) “has just chosen different primary axiologicals from yours.” Now just what can be said about the validity, truth-value, correctness, “rightness” or whatever of one’s primary axiologicals remains unclear, and I have not pretended to settle that. But I urge that the Horrendous Hitler Case doesn’t help us much, because the moral agent in question does *not* differ from the rest of us *only* in his choice of primary axiologicals, but also in a number of other important ways. Having read Hitler’s *Mein Kampf* and his *Table Talk* and his collected *Speeches* and several accounts by friends (e.g., Speer, Doenitz, Ernst “Putzi” Hanfstaengl), I am satisfied that, *quite apart from his axiologicals*, he was a four-star muddle-head, and probably a paranoid schizophrenic to boot. His head was crammed with all manner of nonsense, and he never learned how to think straight about much of anything. As an autodidact, he was a beautiful illustration of Mark Twain’s comment about “self-made men,” that they usually show the horrors of unskilled labor.

But, you may say, what has this to do with the matter of his pure postulates? Well, I’m not clear that he even *had* any, nor that he applied whatever dismal, goofy “pure postulates” he *did* have coherently, nor that he ever reflected on them ERDERVE. As to the factual components of his applied ethic, they are mostly nonsense or, at best, erroneous. So my reply to the positivist is that I will “test” Hitler to see whether he can think non-loony. Having concluded that he can’t, I feel no further impulse to convince him of anything. Why should I? And if he’s in politics, I repeat, the rational steps are to vote against him and then, when he begins to murder people, to dispose of him. An eminently sensible policy for us, is it not?

You may object that I here permit myself a fallacy *ad baculum*, with bloody-minded trimmings besides. But I am only responding to the positivist’s choice of Hitler as example. The question “What do you do with...?” seems to be a question in pragmatics, unless it was carelessly worded. And what I “do with” someone who

disagrees with me about pure postulates depends upon the postulates *and* how the disagreement leads him to act. If Jones is an impressionist in aesthetics and I am a cubist, I do not think it appropriate to kill him. I presume Carnap did not advocate assassination of Brouwer because of the intuitionist/logicist controversy. Nor did Jung try to poison Freud, and so on. When the positivist drags out Hitler to make his point, he does it to emphasize that

- a. Some disagreements are irresoluble.
- b. In ethics, these basic collisions may have great social consequences.

These being his polemic aims in the Hitler example, the positivist cannot complain “*ad baculum!*” when I reply in terms. *If* Hitler really holds a coherent-but-unusual set of primary axiologicals (which I deny), *and* the mass murder of innocent Jews, gypsies, Social Democrats, Poles, etc., is a rationally derivable consequence of his pure postulates (which I also deny), then the pragmatic question becomes, “What do we do with a person who has uncontrolled power and whose ethical postulates are wrong [*sic!*] and who, as a result, proceeds to murder millions of innocent people?” The answer “We have to kill him first” is no more bloody-minded than the policeman’s duty to shoot an armed robber who is shooting at a storekeeper. And it is not an *ad baculum* fallacy because the question itself is one in pragmatics, although intended as a dramatization of the positivist-emoivist thesis. The positivist is not entitled to get a polemic advantage out of Hitler’s evil craziness (“He can’t be persuaded”) plus Hitler’s terrible power (“He can order the Final Solution”), and yet expect us to deal with the pragmatics as if the disagreement were one about the gustatory merits of chocolate versus vanilla ice cream.

10. What directions of inquiry are likely to be fruitful in the next phase of post-positivist ethics?

My answer: First, we need to develop a richer, more formal, and more quantified ethical object language as a suitable subject matter for meta-ethical examination.

Second, the expressions in this language must be analyzed into their “purely axiological” and their “mixed (axiological-psychosocial)” components.

Third, psychological research on ERDERVE convergence should be carried out to ascertain the amount of persisting disagreement about the primary axiological weights in the V-matrix among “sophisticated” person.

Fourth, we should estimate the quantitative impact of such irresoluble pure postulate disagreements upon individual and social choice, given the calculable errors and information gaps in our empirical means-end contingencies. One hopes (and I anticipate) that for a wide range of circumstances, the social choice differences will be small and “contractually compromisable” within the social group’s constitutional framework. There is no reason why we should insist upon universal punctiform

consensus, which we cannot achieve in non-ethical domains either, yet science goes on (and “works”!).

Fifth, psychometric and psychophysiological research should strive for an integrated theory of human satisfaction, and in particular, to converge on a “hedonic calculus” where—despite Pareto and Co.—a psychometrics of interpersonal cardinal utility comparison would be realized.

Sixth, meta-ethics should look to developments in general logic and epistemology, especially the concept *truth* in relation to pure postulation, correspondence, and coherence.

Seventh, the concept *obligation* should be accepted as a primitive, abandoning fruitless efforts to reduce, define, or explain it psychologically (“pro-attitudes,” “commands,” and the like). The emotional-motivational correlates of a person’s cognizing an *ought*-statement should be studied as a branch of semiotic, but not in the hope of cooking up a “definition” of moral obligation in terms of non-moral descriptors.

Eighth, the search for *one single ethical postulate* should be abandoned in favor of explicit recognition that Prichard, Ross and Co. are right about there being multiple moral obligations.

Ninth, the justificationist program should be repudiated, making Hume’s Guillotine harmless.

Tenth, the two-way control between ethical postulates and theorems (whose application to concrete cases “satisfies us”) should be seen as not viciously circular.

Eleventh, the concoction of horrible hypos which yield counter-intuitive results should not be taken as refutation of an ethic, but merely as posing a problem. Ethics should not be held to a higher standard than science, where all theories have troubles at all times.

Twelfth, moral philosophers should become proficient in the technical tools of psychometrics, reinforcement schedules, drive theory, econometrics, and decision theory, especially with an aim to see how much quantitative value-theory can be developed from a combination of *qualitative* pure postulates plus *qualitative* preference rules plus *quantitative* psychosociology.

Finally, “ordinary language” should be relegated, if not to the dustbin, to the minor rôle of suggesting things to look at or worry about; it should certainly not be relied on as a tool of analysis nor, God forbid, as a criterion of anything.

[No list of references was included, because Herbert Feigl would have been familiar with all articles or books alluded to in the memo.—LJY]