How to Derive Morality from Hume's Maxim

Gerald L. Hull, 23 March 1997
ghull@stny.rr.com
Creative Focus
Box 580
Chenango Bridge, NY 13745-0580

[Note that footnotes have been incorporated into the text within square brackets.]

I. Introduction

The argument that follows has a certain air of prestidigitation about it. I attempt to show that, given a couple of innocent-seeming suppositions, it is possible to derive a positive and complete theory of normative ethics from the Humean maxim "You can't get ought from is." This seems, of course, absurd. If the reasoning isn't completely unhinged, you may be sure, the trick has to lie in those "innocent-seeming" props. And, in fact, you are right. But every argument has to begin somewhere, and, however questionable, those suppositions just don't seem to harbor serious normative import.

The disingenuous suppositions are two:

    (1) preferences can be made scientifically respectable and
    empirically quantifiable; and

    (2) morality can be defined as a universally valid means of
    determining the relative weight of different person's
    preferences.

The first is debatable and the second a trifle odd, to say the least. Yet, even if they are both false or otherwise unacceptable, it still should be noteworthy -- as a matter of metaphysical amusement, if nothing else -- that there are *any* suppositions, short of flagrant contradiction (e.g., S & ~S, therefore T), under which the Humean maxim entails a positive theory of ethics. For the import of that maxim appears to be entirely negative; and, as the old Latin slogan proclaims, *ex nihilo nihil*: you can't get something from nothing. But it may not be 100% reliable. Sometimes the absence of something is truly significant, as in the Sherlock Holmes tale of the dog that didn't bark. ["Silver Blaze," in *The Memoirs of Sherlock Holmes*.] Morality, I claim, involves one of those magical instances in which it seems we can actually get something from nothing.

How do suppositions (1) and (2) allow us to derive morality from Hume's maxim? Let's examine them one by one, and that maxim as well, before attempting to put them all

together.

II. Hume's Maxim

First, Hume's maxim:

> (HM) You can't get ought from is.

This is intended to encapsulate a strain of ethical skepticism found, not just in David Hume's run through his library, but also in G. E. Moore's differentiation between goodness and natural qualities and R. M. Hare's distinction between description and prescription. And no doubt elsewhere. [Of course, Hume never says HM in so many words; cf. *Treatise* (Oxford), pp, 469-70.]

HM is often taken to mean

> (HM1) You can't infer 'P ought to be case' from 'P is the case'.

This is clearly true, but, I believe, largely irrelevant. The defense of morality does not usually take such a conspicuously invalid form. (However, exceptions abound: for example, those who take an Aristotelian approach to ethics often appear guilty of inferring "X ought to be man's function" from "X is man's function.") More usually, defenders infer 'Q ought to be the case' from 'P is the case', where P and Q are significantly *different* statements. So a less restrictive interpretation of HM would be:

> (HM2) You can't infer ought statements from is statements.

However, this is too broad -- it begs the question of whether there can be moral truths. For, if they exist, we can indeed derive an ought from an is. E.g., given a virtue approach to morality, we can derive "You ought to be honest" from "Honesty is a virtue." (Note we are not trying to infer "Honesty ought to be a virtue" from "Honesty is a virtue.") There is no logical hanky-panky here. The prescription that shows up in the copula ('ought') of the conclusion is contained in the predicate ('virtue') of the premise. To avoid begging the question, therefore, we need to interpret HM more narrowly as

> (HM3) You can't infer ought statements from scientific
> statements.

More simply, it says you can't get morality from science, where 'science' refers specifically to the established disciplines of logic, mathematics and the natural sciences.

However, even thus qualified HM does not appear to bode well for morality. Both *ontologically* and *epistemologically*, science seems to occupy all the viable terrain. Ontologically, science is the paradigm of objectivity: truths that are *universally valid* (hold necessarily regardless of what anyone may feel or believe). It seems that science describes everything that is objectively real, so there is nothing left over for morality. Epistemologically, there are just two kinds of generally accredited procedures: empirical and conceptual. Empiricism gives us the natural sciences, and conceptuality gives us logic and mathematics. Again, there seems nothing left over for morality. Fortunately for

ethics, neither of these arguments is conclusive.

From an ontological perspective, it may be granted that there is nothing in this world that science does not describe. However, this does not entail that logic, mathematics and natural science *exhaustively* describe the world. An analogy may help. When we speak of the "world" in this context we are speaking literally of *everything*: the infinite sum of all that ever was or ever will be. And when dealing with infinities, our everyday intuitions sometimes fail. In mathematics, for example, it is possible to map a proper subset of the natural numbers, say those evenly divisible by 2, onto all the numbers. Thus the function $f(x) = 2x$ undermines our intuitive assumption that, if one set has more members than another, they cannot be put into 1-to-1 correspondence. Even though, therefore, the even numbers *map onto* all the integers, they do not *constitute* all the integers. Similarly, even though science may map onto (describe) everything in the world, it need not exhaust what may be truly said of those things. There is still room for morality in the metaphysical fissures between logic, mathematics and the natural sciences.

Epistemologically speaking, it is an oversimplification to distinguish natural science, on the one hand, and logic and mathematics, on the other, according to the types of epistemic warrant they receive. Remove the conceptual structure from any natural science -- say, Darwinian theory from biology -- and all you have left is a pile of suggestive data. All sciences have both a theoretical aspect and an applied aspect. Furthermore, while it is undeniable that in some general sense "relations of ideas" and "matters of fact" (to use Hume's phraseology) are the sole warrants for the legitimacy of the established sciences, this places no constraint on the number of distinct disciplines that may thereby achieve respectability. For, once upon a time, physics was as poorly founded as ethics is today -- that is to say, it was considered just another part of philosophy. Isaac Newton's great work was entitled "Mathematical Principles of Natural Philosophy" (*Philosophiae Naturalis Principia Mathematica*). But that very work provided a conceptual and empirical foundation that conferred upon physics, for the first time, the legitimacy of an independent and self-sufficient discipline. Who is to say this cannot happen for morality as well? The present essay is, hopefully, a step in that direction. (Whether this will make morality part of an expanded conception of "science," or something distinct, is, I submit, an unimportant terminological issue.)

Finally, the fact that ethics is separate from (current) science doesn't mean that moral characteristics are completely unrelated to scientific characteristics. Natural science is separate from mathematics, but this does not mean that scientific statements cannot have numbers in them. So morality does not entail the existence of "queer facts," moral properties somehow unhinged from the rest of reality. [J. L. Mackie, *Ethics: Inventing Right and Wrong* (Penguin), pushes this criticism; but even Moore accepts the existence of *synthetic* moral truths, *Principia Ethica* (Cambridge, U. K.).] Consequently, the qualified version (HM3) of Hume's maxim – "You can't get morality from science" -- does not commit us *a priori* to non-objectivity, non-reality, non-cognitivism, &c. But it remains to be seen how a separate morality could be universally valid, let alone how it could be derived from HM.

III. Empirical Preferences

Next, we have supposition (1):

```
preferences can be made scientifically respectable and
empirically quantifiable.
```

This implies that, in contrast to what HM claims about morality, preferences (and at least their *intra*personal comparisons; see below) *are* a matter for science.

The word 'preference' will be taken to refer to the common component of a rich trove of notions concerning what *moves* people to action. A short list: desire, motive, inclination, feeling, purpose, need, sentiment, want, relish, wish, hope, passion, itch, impulse, lust, interest, urge, hankering, hunger, longing, thirst, like, &c. Similarly, on the negative side: fear, hatred, aversion, antipathy, dislike, distaste, &c. Since we want preferences to be as empirical as possible, we put the proof in the pudding -- the ultimate test that a preference exists is that someone is actually *moved* one way or another, to choose and (better yet) to do something. [Note that the rational "interest," with which Immanual Kant accommodates the possibility of people abiding by his Categorical Imperative, counts as a preference on this approach.]

It cannot be denied that the realm of preference is riddled with land mines. Theoretically, it is implicated in a significant array of long-standing and/or fundamental conceptual difficulties: the Problem of Other Minds, the Mind-Body Problem, the Problem of Private Language, and so forth. Practically, the task of determining exactly what a person prefers or intends on any particular occasion can often be insuperable, as any devotee of Court TV can attest. This is because of the natural ambiguity of the circumstances surrounding human actions, and complicating and obfuscating factors like self-ignorance, self-deception and moral weakness, not to mention our skill and proclivity for deceiving others. This has led some philosophers to abandon preference altogether, castigating it -- along with things like thought, belief, intention, choice, action, &c. -- as part of a prescientific "folk psychology" that a more modern psychology (or neural science or computer science or whatever) will ultimately replace.

Still, preferences are conventionally assumed not only in ethics but in scientific disciplines concerned with human behavior (sociology, anthropology, &c.), and are so ingrained in everyday life that it's unlikely any skeptic fails to use them, say, in non-philosophical encounters with law-enforcement representatives. Indisputably, they are unavoidable in moral debate -- if you don't believe that people choose and act, it doesn't make much sense to argue about right and wrong. Consequently, a case can be made for accepting the notion of preference. But, to avoid getting bogged down in all these issues, I will here largely *assume* that preferences can be made conceptually respectable.

To aid and abet that respectability, it will be useful to adopt certain strategies with respect to their empirical measurement. First, as mentioned above, we will *always* infer, from the fact that P chooses or does A, that P prefers A, at least under the circumstances as P understands them. Second, if P considers both A and B, and P chooses or does A, we will *always* infer that P prefers A to B (again, under the circumstances). This obviously enhances the empirical determination of preferences and the measurement of their

relative strength for some given individual (that is to say, in *intra*personal comparisons), but at some expense. For it dismisses the traditional view that people have something called the "will" that allows them to choose & act against their strongest preferences. This issue has been much disputed, providing a good strategic rationale for slighting it here. We resist the "will" in order to render issues about the relative weights of a person's different preferences more empirical.

In assuming it is possible to measure the relative weight of person's preferences, we have assumed we can use "ratio scale" numbers in representing their values. Technically speaking, this implies that we can speak of *how much* more important one preference is than another, and can identify a *natural zero* (i.e., indifference) in our weightings of them. This legitimizes the use of fractions and ratios in the representation of (the weights of) preferences, e.g., 'N/M' and 'N:M'. The importance of this will soon become clear. It should be emphasized, however, that numerical representation of the relative value of independent preferences will often be extremely imprecise and fuzzy.

The empiricality of preference is not altogether congenial to conventional approaches to morality. For example, it opens the possibility that a person may have many independent preferences, and that preferences may differ arbitrarily from one person to another. In particular, the importance of specific preferences may vary greatly among individuals: what's most important to me may not be what's most important to you. These possibilities are fatal to many approaches: for instance, those that assume that there is an underlying, common element to all preferences. (Some Utilitarians claim that everyone is ultimately after happiness, that is to say, pleasure *sans* pain.) That assumption may turn out false or empty if the content of preferences is regarded as entirely empirical. In our lives we may find many logically independent preferences with no common ingredient. Similarly, moral sense and evolution-based arguments often claim everybody has certain preferences in common, which are taken to define moral attitudes. But since, in fact, genes differ from one person to the next, you may have a different sensibility than I, or a stronger or weaker, or even none at all (i.e., be a sociopath). Can the moral sense be binding on someone who doesn't have it? Finally, any attempt to derive a social welfare function from the welfare functions of individuals is undermined (Arrow's Theorem) when preferences can vary arbitrarily from one person to the next. So rather than supporting morality, the insistence on the empirical significance of preference tends to subvert it.

IV. Interpersonal Comparisons

Last, we have supposition (2):

```
morality can be defined as a universally valid means of
determining the relative weights of different persons'
preferences.
```

Many different interpretations of morality have been introduced during the history of philosophy, and while this one may seem curious, simplistic and/or deviant, it is hardly the strangest ever proposed. The remainder of this essay will provide ample opportunity

for determining just how productive this definition may be. In one respect this characterization does represent only half a loaf. Traditionally, ethics is conceived as something that can be self- as well as other-directed. For example, the Seven Deadly Sins include individual failings like Sloth and Gluttony as well as social defects like Lust and Avarice. Accordingly, Kant distinguished between *duties to oneself* and *duties to others*. Still, the social dimension of morality is hardly unimportant, and if some success can be sustained by setting our sights there, few should complain. And through universal validity we may hope to attain an ethics that is both objective and real. This, note, is precisely the hope generally thought to be dashed forever by Hume's maxim.

The phrase 'relative weights of different person's preferences' is a bit vague. Fortunately, earlier assumptions -- in particular, that preferences may be quantified using a rough ratio scale -- allow us to make its significance much more precise and concise. [Note that when quantifying preferences I will speak variously of their "weight," "importance" and "value", and sometimes, as above, simply speak of "preferences." Of course, the content of a preference should not be confused with its relative importance for some individual; context should make it clear, in any instance, what meaning is intended.] For any given individual, there will be some preference that is the most important for that individual. What preference this is will be an entirely empirical question, so we may not assume that what's most important for one person is the same for another. (Actually, nothing prevents more than one preference being most important for some person -- e.g., someone may honestly be unable to choose between the relative advantages of family life and a life devoted to the ballet -- though, fortunately, such dilemmas tend to be rare. For purposes of simplicity, I will henceforth assume there is just a single "most important" for any given individual.)

That which is most important for a person provides a convenient unit measure for gauging the relative importance of any of that person's preferences. They all can be expressed as fractions with what is most important as the common denominator. Consequently, a person's greatest preference provides an *evaluational* or *preferential yardstick* for that person, embodying that person's preferential values. Insofar as people's values differ, estimation according to their preferential yardsticks will also differ. (And yes, with "weights" and "yardsticks" I am mixing my metaphors.) Since all of P's preferences can be evaluated in terms of P's personal yardstick, and the same for Q, all we need to know to evaluate the relative importance of any of P's and Q's preferences is the ratio of their respective yardsticks. Since we have supposed that morality can be characterized in terms of the relative importance of different person's preferences, it follows that the fundamental issue of morality, on this approach, boils down to the ratio of person's unit measures, their preferential yardsticks.

Consequently, whether there can be an objective, real ethics turns on whether we can find a universally valid means of determining the ratio of different persons' evaluational yardsticks. Of course, we are not simply interested in what relative weight is *in fact* ascribed to different persons' preferences on any particular occasion, as implicit in their choices and actions, but what weight *ought to be* ascribed to them. The former issue can be addressed by science, given our empirical construal of preferences; the latter, in accordance with Hume's maxim, can only be addressed by morality.

V. Ratios and Conflicts

Why frame morality in terms of ratios of preferential yardsticks? At best, this may seem to be no more than a peculiarly roundabout approach to Utilitarianism. [Or a pathetic masculine metaphor.] But in fact it is quite different, and provides an efficient yet unrestrictive approach to the resolution of interpersonal conflict. The great normative defect of Utilitarianism is that it washes all preferences clean of personal ownership, as if they belonged indifferently to a single, all-encompassing entity -- the great social Leviathan. But this implies that in principle there is nothing wrong with one segment of the populace continually suffering harm as long as everyone else reaps commensurately greater benefits. But this is absurd, as members of the afflicted segment would no doubt make clear: "Why should we suffer for the sake of everyone else?"

Instead of regarding the weights of different person's preferences simply as fodder for some summing function or as indices of relative importance in some decision matrix, the present approach interprets them as *rights* to satisfaction. Or, inversely, if your preference is deontology, they represent *duties*: for what I have the right to demand, you have the duty to provide, and *vice versa*. This plays out most transparently in the context of distributive justice. If we are apportioning some divisible good, and A's and B's preferences in that regard weigh M and N respectively -- and all other things are equal! -- then M:N determines the relative proportions they should be accorded. [Shades of Karl Marx! But the valorous entrepreneur should not abandon hope; see section XI.] But often things are not equal. More generally, in my pursuit of some preference I do not have the right to collaterally frustrate your preferences. If crossover interference cannot be avoided, then some *quid pro quo* agreement must be arranged that adequately reflects the relative amounts of our respective preferential involvements.

How this works out in normative detail will be the focus of a later section. At the present time, however, we still don't know how to weigh the preferences of different people: how to *calibrate* their respective preferential yardsticks. But if we could do that, it should be evident that this preferences-as-rights approach would provide a means for judging whether any given option -- action, proposal, alternative, institution, strategy or whatever -- is right or wrong. It is right if it properly reflects and respects the preferences of the individuals involved, and wrong if it doesn't. Assuming for the moment that the details of this approach can plausibly be elaborated, it shares one great virtue of Utilitarianism. That is, it provides a *systematic* means of assessing the moral status of options: it provides a consistent procedure for assessing the options involved in any dispute between individuals. Indeed, insofar as preferences have been rendered empirical, it provides a procedure that is not just decidable but complete, so far as possible parties to conflicts are concerned. So if we can only find a universally valid means of determining the ratios of persons' evaluational yardsticks, *we will have a universally valid -- and consistent, decidable and complete -- means of resolving interpersonal conflicts.*

However, there are some philosophers who feel that this kind of systematicity is out of place in morality [cf. Hare, *Moral Thinking* (Oxford), p. 212; Mary Midgley, *Can't we make moral judgements?* (St. Martin's), p. 143]. They argue, on various grounds, that an

effective procedure for telling right from wrong is unavailable, inappropriate and/or unwanted. Let me suggest that their position fails to appreciate what such a procedure could, and more importantly could not, do. It is one thing to determine what is *right*; it is quite another to determine what is *absolutely best*. While the present approach suggests a way of accomplishing the former, it seems clear that the latter will never be possible. This is because, in any given circumstance, there are an indefinite number of ways of dividing up the possible alternatives: depending upon one's interests, upon assumptions concerning what the laws of physics, &c. may allow, upon what goals are at stake, upon the capacities of the agents involved, and so forth. Indeed, if only because no person in any choice circumstance can ever know all that could ever be known about that circumstance, the ways of dividing up possible alternatives are not enumerable.

This implies that, while one may be able judge the rightness of various suggested solutions -- and thereby assess which is *relatively best* -- there is no way to automatically generate the absolutely best solution to a particular moral problem. This is quite important, if for no other reason that, from the time of Aristotle on, philosophical moralists have concerned themselves with the question: "What is the best life for a person P to lead?" We have just shown that this question cannot be answered; for, in effect, it reduces to asking, time after time, an unanswerable question, "What is the absolutely best option for P at this time?" Consequently, the existence of an effective procedure for telling right from wrong in no way eliminates the challenge, the doubt, the hard thinking and the creativity that morality demands of us. There is an illuminating analogy here with the world of mathematics. In arithmetic, we know from the work of Kurt Goedel, Alan Turing and others that there is no effective procedure for determining whether a given statement is a theorem, though there is such a procedure for deciding whether any purported proof of a theorem is valid. Analogously, in ethics there is no effective procedure for generating the absolutely best option, though there is one for determining whether any given option is morally correct.

As remarked, this does not mean that we cannot determine the relatively best of a specific (circumscribed) set of options, and surely on many occasions that would be enough to satisfy us. Undoubtedly, many of the breakthrough instances of human moral genius have involved discerning new, previously overlooked solutions to moral dilemmas. But, on a day-to-day basis, morality will nonetheless be a great boon insofar as it allows us to determine whether particular options are right or wrong. Here the difference between the present approach and Utilitarianism should become conspicuous, for it opens the possibility that, under some circumstances, what is (relatively) best may not be entirely right. That is, we need to heed the old saw about having to choose the lesser of two evils. For sometimes, as when the madman forces us to select between boiling one baby alive and boiling ten babies alive, neither of the available options is "right." Sometimes, surely, circumstances are such that *all possible* alternatives are wrong in some respect.

VI. Despicable Preferences

The usefulness of defining morality in terms of the relation between different persons' preferential yardsticks shows itself when we confront what can be called the Dilemma of

the Despicable Preference. Frustrating any attempt to found morality upon preferences is the fact that some of them are, well, *despicable*: for example, things like "sadism, envy, resentment, and malice" [John Harsanyi, "Morality and the theory of rational behavior," in *Utilitarianism and beyond* (Cambridge, U. K.), ed. Amarta Sen and Bernard Williams]. This poses the dilemma that, either we accept such preferences (horn #1), at the risk of rendering the resulting ethics normatively intolerable; or we reject them out of hand (horn #2), thereby presupposing what counts as ethical right from the outset. Utilitarians, confronted by this dilemma, typically grasp horn #2 [cf. Sen & Williams, "Introduction," *Utilitarianism and beyond*]. Without going into those difficulties, it should be clear that it would be fatal to the present approach: it would destroy the empirical foundation we have labored so hard to establish. For now the weight of a person's preferences would not be determined as much by what that person actually feels, and therewith intends and acts, as by what some person sitting in an armchair somewhere regards as decent and tolerable. Instead of laying a foundation upon which morality may be later erected, we have smuggled it in with the bricks and mortar.

The problem posed by despicable preferences is not that they conflict as a *matter of fact* with those of other people. This would be no different than the run-of-the-mill conflicts of interest. Rather, it is *logically* the case that malice, for example, cannot be satisfied without the victim suffering dissatisfaction (to put it mildly). They represent cases where preferences have as their very object the denial of the preferences of others. How can we exclude them without invalidating our whole approach?

The answer lies in the definition of morality we have adopted. Since it concerns the ratio of different persons' preferences, any preference that presupposes some such relative estimation *begs the question*. That is, the goal of this enterprise is to establish some universally valid ground for comparing evaluational yardsticks. So until that ground is established, we must on principle exclude any preference that would prejudge the question. This is analogous to the standard proviso in the sentential calculus that precludes interpreting the primitive variables as logically complex statements. You need to assume that certain truth value assignments are all equally possible in order to define the various sentential operators: "and," "or," "not," &c. That assumption is likely to be false if the basic variables are logically complex. In other words, it is the point of the calculus to *explicate* all relevant logical relationships; smuggling them in *via* the primitive variables would necessarily frustrate that point.

Similarly, since we have defined the search for morality as a search for a universally valid means of determining the relative importance of different persons' preferences, we must start with preferences that don't preempt that search. We cannot include malice, in other words, because that would allow that it is desirable for someone to experience something they consider undesirable, thereby prejudging the relative importance of the preferences of *malicer* and *malicee*. To begin with, therefore, only "logically sanitary" preferences can be admitted for consideration. This excludes benevolent as well as malevolent impulses. Only when we have determined what is moral and what is not can we evaluate more complicated preferences that presuppose the relative importance of different persons' preferences. Then we will be able to say if malice is moral or immoral (presumably the latter). To employ the standard term of art, our search for morality must

be restricted to *first order* preferences: i.e., preferences that do not have other preferences as their objects.

[The conclusion just reached actually goes beyond the argument presented. A complete treatment would have to consider intrapersonal as well as interpersonal higher order preferences. That is, to start with, we also want to exclude cases wherein P has preferences concerning P's other preferences, e.g. the smoker who wants to quit. This is not really our concern insofar as we have restricted our attention to the social, other-directed dimension of morality. But, for the sake of simplicity, it will be convenient to assume that any higher order preference is initially inadmissible.]

VII. Hume's Maxim Reprised

Despite assorted difficulties, the assumptions that preferences are scientific, and that morality concerns their relative weight, at least in the context of moral debate, seem relatively innocent and unpretentious. How, now, do they inseminate morality in the unwitting womb of Hume? We have supposed that we can derive *intra*personal comparisons from science but not *inter*personal comparisons. In particular, HM entails we cannot get inter- from intra-personal comparisons. Science (psychology, sociology, anthropology, &c.) may provide an empirical basis for measuring the relative strength of P's preferences to other preferences of P, but not to those of any other person, say Q.

To appreciate the force of HM, suppose that at some later time brain science identifies the neurological substrate of preference. In particular, suppose that there is some parameter, K (certain nerve-firing frequencies, or neurochemical levels, or whatever), that strongly correlates with the relative strengths of persons' preferences as revealed by their behavior as well as their subjective self-assessments while in contemplative repose. Further, suppose that it is discovered that the K-values of some persons (or, worse, of some "racial" genetic sub-types) are higher, across the board, than those of others. The question is: would this provide an entirely scientific determination of the relative importance of different persons' preferential yardsticks? The answer is: No! For no amount of scientific measurement would tell us how to interpret those K-values. We could still argue either (1) A deserves more than B because A feels more strongly (has higher K-values), or (2) B deserves more than A in order to bring B's feelings up to A's level, or (3) A and B deserve the same because K-values should be treated as relative to the individual (i.e., to their evaluational yardsticks).

Again, we butt our heads up against an impassable barrier -- morality cannot be derived from science. But, of course, we cannot (meaningfully) get morality from itself: that would beg the question. Nor does there appear to be some third party -- religion, metaphysics, space-alien super-science, &c. -- from which we might somehow derive morality. For any such would simply raise again the very same question raised by morality: insofar it is distinct from science, what constitutes its epistemological and ontological legitimacy? For this reason, most approaches to morality regard Hume's maxim as a serious threat, and make special efforts to diminish or deflect its significance. Instead, I say, let's take it to the hilt!

VIII. Something from Nothing

Insofar as preferences are empirical, it is an empirical question whether P's choices or actions affect Q's preferences, and vice versa. Since human beings aren't typically isolated from one other, this will often be the case. So we can raise this question: in P's choices, what is the relative importance P accords Q's preferences? Now P may have affected Q's preferences unknowingly or carelessly or whatever, but still P, in every choice or act, willy-nilly presupposes some ratio R representing the relative importance of P's and Q's preferential yardsticks. For example, if P steals candy from baby Q, among other things P is implying, "P's hankering for sweets is more important than Q's." So what can P possibly say in justification of some particular ratio, R, that he has assumed between P's and Q's yardsticks? This, note, asks a moral question: what *ought to be* the ratio? And it is a question that Q will raise for sure, assuming Q is rational and able, when P tries to tip things in P's favor.

Let us take first the infinite number of cases in which R = N:M, where M and N are two different numbers (M <> N). We know from Hume's maxim that there is no justification for any such arbitrary ratio. There is no fact of logic, mathematics or natural science that allows us to deduce it. Nor is there some other legitimate source, outside of science, to which we can appeal. So we know that, in every case of this kind, the assumption of the ratio R is irrational. According to HM, no reason can be given for it. Next, let us take the singular case in which ratio R = N:N (which, of course, reduces to 1:1). In this case, and this case alone, P can give a reason! P can cite HM in support, to wit: "There's no reason to value the one more than the other." Consequently, once we raise the question of the relative weight of different person's preferences, there is one and only one possible rational answer: their yardsticks should be regarded as equal.

Seems like sleight of hand, no? Take the same argument from a slightly broader perspective. Suppose we have X number of persons debating whether or not to take a particular course of action. That course of action will affect the preferences of the X persons in a certain empirically determinable ratio, R = Y1 : Y2 : ... : YX. People show a tendency toward selfishness, in that they are attracted to courses of action in which their own preferences receive generous attention, i.e., their allotment of benefits is as large as possible. But any partiality that one person receives can only come out of the allotment for someone else. The perspective of ratios makes this a zero-sum confrontation. Now suppose the offended party asks for a reason -- some kind of justification -- for this comparative ill-treatment. We know from HM that no valid reason will ever be forthcoming. So unequal treatment is literally *irrational*. However, a person's justification for complaint vanishes at the equal point: any farther, and others acquire the status of offended party. Consequently, equality of preferential yardsticks represents a singularity -- an equilibrium point -- the only ratio of preference satisfaction (with respect to some course of action) to which none of the X persons can have a valid gripe. In other words, generalizing, it is the only *universally valid* ratio.

[Note the difference between this and John Rawls' Original Position argument, *A Theory of Justice* (Harvard). I am not claiming that this conjectural debate represents a

fundamental procedure for determining justice. Rather, it reveals a logical fact about the nature of morality: the only rationally defensible position is to regard everybody's evaluational yardsticks as equal.]

IX. The Trick Revealed

How have we gotten away with getting something from nothing? 'Tis time to 'fess up: the argument here has indeed involved sleight of hand. The old Latin saying was correct: you can't get something from nothing. The trick lies in very special circumstances that permit the derivation of something of overriding importance from, not nothing, but the merest logical triviality. In effect, Hume's maxim clears the field -- since science is completely separate, nothing in it has the least consequence for morality -- thus creating a *tabula rasa*. But regardless of this separatism, morality promises to be terribly useful if it can provide an effective and universally valid means for resolving interpersonal conflicts. It is the assumption of ratio scale values for preferences, from supposition (1), plus the need to apply them in a universally valid way to person's evaluational yardsticks, from supposition (2), that does the trick. By framing interpersonal relations in terms of the relative weight of different person's yardsticks, that *ratio* becomes crucial to morality. If we have it, we can evaluate actions, options, &c. in terms of satisfactions (or dissatisfactions) of different persons' preferences, and thereby effectively tell right from wrong.

It is a trivial fact about the logical -- or, if you will, *logico-mathematical* -- structure of ratios that puts a smudge on the *tabula rasa* created by HM. It is the evident fact that ratio N:N is less arbitrary than any other ratio N:M, where M $<>$ N. This is simply because we need one less number to specify it; N:N has one less degree of freedom because we don't have to account for M. And, obviously, the value of N:N is invariant with respect to N, so neither is it arbitrary on that score. Let me reemphasize that this, in itself, is an insignificant point of logic (& math), and of absolutely no ethical significance (in accordance with HM). But morality requires, we have supposed, that we determine the relative weight of different persons' preferential yardsticks. This requirement explosively transforms that logical triviality, turning the goose girl into a fairy princess. Since we need a ratio, and moreover one that is universally valid, we are left with the one ratio that is least arbitrary, *viz.* N:N (aka 1:1). Any other will be less than universally valid because it raises a question that, by HM, is unanswerable: What justifies M $<>$ N? Someone in the universe will have a legitimate ground for complaint. So, in the absence of any other constraint, the consideration of logical simplicity becomes determinative. *If we must have a universally valid ratio*, N:N is the only viable choice.

And why does M cause trouble, when M $<>$ N? It violates the fundamental requirement of *rationality* itself. To be rational is to provide reasons. One need not go overboard in affection for ratiocination to admit that it enables some pretty amazing feats. If you can provide valid reasons, things are no longer arbitrary. You can discriminate intelligently. This is as true in mathematics and physics (&c.) as it is in morality. If A and B are different, an adequate (rational) discipline must explain that difference. If one projectile lands on target and a second goes wildly astray, physics is obliged to provide a reason. A

physics that could not provide this answer, in principle, would to that extent be incomplete and irrational ("Duh, it just happens that way!"). Rationality requires that there be no difference without a difference (*pace* QM). The same holds *a fortiori* for rational morality. If P and Q are to be treated differently -- if their yardsticks are not weighted the same -- there arises the need for an explanation of that difference. In virtue of Hume's maxim, any morality that assumes a non-equitable ratio N:M between P's and Q's preferential yardsticks is thereby arbitrary and irrational. [*Contra* J. R. Lucas, *Justice* (Oxford).]

It helps to look at the situation from a human perspective. Because of the logic, any proposed ratio of the form N:M (N <> M) is rationally objectionable, and therefore objectionable to everybody -- universally invalid. But humans, it has been remarked, tend to be somewhat selfish. We are less likely to complain when we are the *beneficiaries*. We tend to be simplistic game theorists: we are much attracted to the larger immediate payoff. But we are not nearly as comfortable with inequality when we are on the losing side. So for any group of persons involved in some controversy, although everyone has a valid complaint against inequitable ratios, it is those on the short end of the yardstick that we expect to raise the objection, "For what reason *ought I* be getting less?" No valid answer will be forthcoming: the consequence of Hume's maxim, "You can't derive ought statements from scientific statements." Of all possible ratios, only N:N provides no equivalent basis for legitimate complaint. If I try, I will be told, "What's your beef? You're not getting less -- you're getting the same treatment as anybody else." No one can attempt to get more than equality without giving someone else both basis and motivation for complaint. Therefore, only N:N (aka 1:1) provides the universal validity required by morality. This is true regardless of what we may think or want. Ignorance provides no escape. Insofar as we are selfish, it is equally unacceptable to all of us. But insofar as we are rational, N:N is a necessary, objective feature of moral reality.

X. Ought from Is?

But how have we come to this pass? We have apparently derived a fundamental principle of morality, to wit,

> (FP) All person's preferential yardsticks ought to be regarded as equal.

How can this have happened without violation of the major premise underlying the entire argument, Hume's maxim: "Moral statements cannot be derived from scientific statements"?

One possible answer to this question is superficially attractive but revealingly mistaken. Consider HM itself: what kind of truth or fact is it? Clearly not a truth of science, that is, of logic, mathematics or natural science. Science cannot state what is not in science without drawing that thing, contradictorily, into science. To see and say that morality is not in science is to assume a perspective that steps outside of them both. Consequently, insofar as we have derived a normative moral view from HM, we have not violated it

because HM is not itself a scientific statement in the required sense. This argument, albeit temptingly ingenious, fails because we have confessed that the morality we have derived doesn't come simply from Hume's maxim, and because HM is not a moral statement either, for the same reason it isn't scientific. It is outside of both: it is *metaphysical*, for lack of a better category. So the question persists: Where does morality come from? One thing that the revelation of the non-scientific nature of Hume's maxim does show, however, is that there are some truths that aren't truths of science! As a consequence, the idea that there may be moral truths becomes that much more palatable.

What we have derived can be more explicitly spelt out as:

> (FP1) *If you want universal validity*, all person's preferential yardsticks ought to be regarded as equal.

So we have not derived an ought from an is, but instead have derived a moral ought from a non-moral or pre-moral ought. If (pre-moral ought) you feel you should assume a stance that is universally valid, then (moral ought) you should take the preferential yardsticks of others to be equal to your own. In other words, given the desirability of a universally valid ratio, equality is the only way to go. [Regrettably, there isn't enough space (or computer memory) to go into all the implications this has *vis-à-vis* Kant's distinction between Hypothetical and Categorical Imperatives.]

Why desire an objective morality? There are, in fact, many attractive aspects to a universally valid means of assessing right and wrong. It provides a systematic framework for conflict resolution that *applies to* any and all participants, and *can be applied by* any and all participants. Kant skewed moral philosophy with his emphasis on *obligation*. It should be obvious that the other face of obligation is *empowerment*: in a closed social system, what comes out of one person's pocket must wind up in someone else's. Ethics allows you to decide what to do in a fair way that can be justified to anyone. Regardless of your own commitment to morality, it enables you to assess others to determine whether they are threats, i.e., aren't likely to be fair with you. From a society-wide perspective, you can judge the conduct of persons with respect to suitability for reward or punishment. More generally, you will be able to frame laws and institutions so that the society as a whole will be more likely to prosper (i.e., so there will be more inducement to cooperation and less disgruntlement). And, for what it's worth, treating person's preferences as equal will lead to success at the Prisoner's Dilemma. In short, there are a number of advantages in having an objective morality, both in knowing right from wrong, and in doing what is right.

Now, of course, some people don't desire universal validity, at least insofar as their own actions are concerned. They are concerned with number one, or with their limited circle of cronies, and the rest of us can just go hang! So it's important to recognize that, on this approach to morality, there is no "internalist" puissance [*a la* Thomas Nagel, *The Possibility of Altruism* (Oxford)] that somehow compels people to do what is right. Knowledge of morality no more guarantees moral behavior than knowledge of logic entails consistency, or knowledge of science precludes wishful thinking. If anything should be apparent, at this late stage of the game, it's that the advantages of rationality are not equally appreciated by all persons at all times.

The objectivity of the normative rule that all (first order) preferential yardsticks are equal does not consist in its *universally being regarded* as valid. But, of course, neither is science objective in that essentially contingent, demographic sense. Rather, its objectivity consists in certain penalties that must be paid, willy-nilly, by anyone who is ignorant, careless or contemptuous of it. [This is how we rework Kant's necessity.] If you eschew logic, you lose the guarantee of truthfulness in your conclusions. If you abandon mathematics, you forego any correspondence between your sums and the quantities they are supposed to represent. If you disregard natural science, you will not prevent causes from being followed by their effects (unlike the world of cartoons, where gravity awaits recognition of absence-of-support before it goes to work). In just the same way, people who neglect the constraint of morality in their choices and actions necessarily abandon the ability to justify their behavior to others. Some persons, alas, are not interested in whether their behavior can be justified. Consequently, the rest of us must protect ourselves from them as best we can. Be that as it may, immorality is inherently indefensible. Ethics is objective in that it is independent of what anyone may think or feel. This does not mean that it cannot concern preferences -- else, for the same reason, psychological laws could not be objective. Rather, it means that what is right does not depend on what anyone might believe or prefer to consider right. "That's not fair!" is not an Austinian performative utterance.

XI. Normative Implications

A case has been made for the derivation of an objective normative ethics from Hume's maxim plus a couple of suppositions about the nature of preferences and morality. The argument has been complex and unavoidably cursory in places, leaving plenty of room for principled disagreement by those disinclined to accept the conclusion. Much the same can be said of any attempt to provide a rational basis for morality. Surely it is no accident that the Devil, in folk tales around the world, is typically represented as a very clever, intellectual type who ensnares his victims with webs of words. So let us put all that aside, and ask instead: What kind of normative ethics does this approach entail, and how does it measure up to our customary intuitions about the nature of morality? [Here we catch a glimpse, in our rearview mirror, of the dispute over whether normative ethics can be derived from metaethics.]

Intuitions and feelings, of course, are generally disdained by philosophers. They provide no arguments and, unless they just happen to be shared, provide no basis for common agreement. But, in fact, people across the ages have shared certain general intuitions about right and wrong. And philosophical defenses of particular moral views often take great pains to show how they comport with what people commonly accept. Consequently, it is not at all improper to examine the normative substance of the current approach to see how well it passes that test. Its overall import has been already mentioned:

    (FP) All person's preferential yardsticks ought to be regarded as
    equal.

In other words, it is wrong -- literally unjustifiable -- to treat one person's (first order)

preferences as more important than someone else's. [Note that I sometimes speak generally of 'preferences', instead of more specifically of 'preferential yardsticks'; in a context like this, they amount to the same thing.] But how this works out in particular cases may not be so clear.

Let's start with Kant's archetypal duty to others: the prohibition against lying. It is obvious that there are advantages to having a correct view of the way things are, and disadvantages to having an incorrect view. People are agents: they perform actions in order to promote their preferential goals. For this agency to succeed, correct information about surrounding circumstances and their own capabilities is crucial. False information is often a sure-fire prescription for failure. So, at least in general, a deliberate falsehood can be expected to put those who believe it at a comparative disadvantage in satisfying of their preferences. Therefore, as a rule, lying is wrong.

But is lying *never* justifiable? Famously, Kant argues for the affirmative. Even when someone who wishes to murder our friend demands to know where he is, Kant insists we should never tell a lie ["On a Supposed Right to Lie from Altruistic Motives," *Absolutism and Its Consequentialist Critics* (Rowman and Littlefield), ed. Joram Graf Haber]. This absolutism seems laughable in an age when the declarative sentence is regarded primarily as a tool for representing affairs in a light most beneficial to the declarer. Without endorsing this depreciation of discourse, it is apparent on the present approach that there are occasions -- e.g., when dealing with murderers -- when a lie actually advances the preferences of others. Does Kant really think that his friend, with dying breath, will celebrate Kant's steadfast commitment to veracity? Indeed, as noted earlier, we can agree with Kant that lying is intrinsically wrong but still claim that in some constrained circumstances it is nonetheless justifiable: when it is the *least wrong* of the available alternatives.

How about the preferences of the murderer -- don't they count? As a matter of fact, they don't, for reasons already suggested in the section on despicable preferences. Intending to kill someone involves a higher order preference: to render all of the victim's preferences null and void. Thus it does not get in on the ground level. And, clearly, preferring to negate the preferences of someone else, all things being equal, is an immoral preference insofar as it denies the legitimacy of the intended victim's preferences. *There is no moral weight to immoral preferences.* To give Kant his due, however, it's important to underscore just how rarely lying is justifiable, and just how risky it is to lie on any particular occasion. The future will always be uncertain to some degree, so we can never be sure the overall consequences of our lie will be beneficial to others. Moreover, it is always somewhat presumptive to purport to know the preferences of others. Perhaps our friend is a diehard deontologist, and truly prefers his death to our dishonesty. And humans are invidiously afflicted with self-deception, so that often our pretended sooth-sacrifices for others are actually self-servings in disguise.

To digress for a moment: it may have been noticed that all of the normative pronouncements so far have been accompanied by some form of *ceteris paribus* qualification: "in general," "as a rule," "all things being equal." Does this show that moral judgments are different in principle from those of accepted sciences like physics? I suspect not. Even in physics, when predicting the motion of a body it is necessary to

assume that all of the relevant masses, forces, &c. have been identified. There are two reasons why morality seems different; one is remediable, the other not. First, up till now ethics has had no generally accepted framework for accommodating additional factors, unlike physics where you can simply add new dots, draw new parallelograms, &c. But with the success of systematic efforts like the present, perhaps that disadvantage can be overcome. Second, for the most part we know where and how to look for the kinds of factors that can perturb a physical model. This is not at all the case in morality, which carries action-at-a-distance to an absurd extreme. A secret promise made centuries ago in a land far away can tip the balance from right to wrong in something done today. So while it is possible to make firm moral judgments given limiting assumptions about the factors involved, in real world applications there will always be the nagging possibility that something significant has been overlooked.

Getting back to the point: with respect to lying, therefore, we can see how the present approach differs from Kantian absolutism. In this, I submit, it is closer to our common moral sense than strict deontology, insofar as it makes it clear why lying is bad but still allows exceptions under certain extraordinary conditions.

But the reader may expect that the real weakness lies in the *egalitarianism* of the prescription that everybody's preferences (more precisely, their evaluational yardsticks) be seen as equal in the eyes of ethics. To begin with, does this not imply that the preferences of Charles Manson are just as legitimate as those of Jones Salk? The answer here is simple: their first order preferences *are* equal; their higher order preferences *are not*. It is fundamentally with respect to the latter that these two gentlemen differ. The former, as murderer, denies the legitimacy of the preferences of others; while the latter, as healer, acts to foster and fulfill the preferences of others. The current approach thus allows us to make a principled distinction between the legitimacy of the preferences of vicious and generous individuals, as our ordinary intuitions would require.

What, then, about egalitarianism *per se*? Does not FP imply that everyone should get equal treatment as long as they aren't explicitly immoral? What about the Aesopian ant who labors long and hard while his cousin, the grasshopper, indulges in naps? Here again the present approach allows us to make a principled distinction. For both ant and grasshopper, let us assume, would equally enjoy napping. But then to give them equal shares of the harvest would be to give preferential treatment to the grasshopper, who reaps both nap and grain while the ant gets only grain. It is easy to see, consequently, that the present view supports the common notion that -- all other things being equal -- people's rewards from a successful enterprise should be proportionate to their contributions. [Cf. Herman Melville's ship-shares in *Moby Dick*.]

On the general subject of egalitarianism, let me suggest that the debate has often been miscast. It is not so much over whether people are equal -- almost everyone will allow that people are equal *in some respect* -- as over when they are unequal: over what should be allowed to count as a significant moral difference. Some radical egalitarians seem to suggest that there are never grounds for treating people differently. The approach that has been described here, as we have seen, does not make that indiscriminate mistake. More important, for the resolution of moral conflicts, is that the explicitly quantitative nature of FP allows us to derive, not just *that* people should be treated differently in specific

circumstances, but *how much* differently they should be treated.

A final challenge strikes at our means of interpersonal comparison. What about the sadist whose joy in causing pain is *equal* (according to their respective yardsticks) to some ordinary Joe's aversion to suffering? Must we, therefore, split the difference, and allow the Marquis modest gratification and Joe moderate discomfort? Since the Marquis' sadism clearly involves a higher order preference, we must determine the morality of allowing his satisfaction at the expense of Joe's dissatisfaction. But this would assume the Marquis' preferences are more important than Joe's, which violates the equity of yardsticks. We can now see why the sadist has no claim to pain. It has already been noted that there is a natural zero -- indifference, neutrality -- when it comes to the satisfaction of preferences. Consequently, we must distinguish positive preferences, whose satisfaction yields a net gain (e.g., winning a prize), from negative preferences, whose satisfaction merely avoids a net loss (e.g., avoiding a fine). While partial fulfillment of our sadist's preference would be to his advantage, partial satisfaction of Joe's would still result in his privation. The only way for both to receive the same (signed) benefit -- the only way to treat their preferences as equal -- is for no interaction to occur (no joy and no suffering), just as common moral sense would dictate.

This survey of the normative implications (FP) of the present approach to morality hardly constitutes a rigorous test of a system that is supposed to measure right and wrong for any option or alternative, up to and including the structure of social institutions. But at least it should show that our derivation of ethics is not entirely off the map, so far as customary moral expectations are concerned. In this, I submit, it has a leg up on many approaches whose theoretical sophistication comes at the expense of their practical plausibility. [For example, Peter Singer condones infanticide in *Practical Ethics* (Cambridge, U. K.), while David Gauthier condemns the handicapped to second class citizenship in *Morals by Agreement* (Oxford).]

XII. Conclusion

This brings our little post-modern ethical conjuration to a close. A trick has been performed, then its underlying mechanics revealed. I have seemingly gotten a positive, normative ethics from Hume's maxim, "You can't get an ought from an is." But while that maxim was crucial, the trick relies decisively on additional suppositions about the nature of preferences and the character of morality. In effect, morality comes into existence because *we would like it to exist*: because it alone provides a universally valid basis for determining right and wrong, and because this is something everyone can use. Maybe the real sleight of hand occurs when we change our attitude towards HM. Conventionally, it is regarded by supporters of moral objectivity as a kind of metaphysical vampire, to be avoided at any cost. Instead, I suggest, real progress can be attained only when we embrace it wholeheartedly. Instead of tolling the deathknell of ethics, Hume's maxim celebrates its rising up from the dead. This is not unprecedented. In Greek mythology, the Phoenix is reborn from its ashes. In more recent times, we are advised, when handed a lemon, to make lemonade. Perhaps the best gloss can be taken from prescient remarks by one of this century's most renown philosophers:

A person caught in a philosophical confusion is like a man in a room who wants to get out but doesn't know how. He tries the window but it is too high. He tries the chimney but it is too narrow. And if he would only *turn around*, he would see that the door has been open all the time! [Ludwig Wittgenstein, from Norman Malcolm's *Memoir* (Oxford), p. 51.]