Cognitive Naturalism in Metaethics

LEIGH B. KELLEY
University of Tulsa

So-called "rational motivation theories"[1] in metaethics constitute an interesting and varied category. "Examples...are ideal-observer and qualified-attitude theories and certain versions of contractarianism - cognitivist theories which give an account of moral truth in terms of a class of rationally constrained motivational responses."[2] Indeed, the scope of such theories is often far broader than "moral truth" *per se*, and may include explications of overriding normative judgments, essential requirements of practical rationality, the concept of value (taken in its strongest sense) and a host of longstanding metaethical problems.

A somewhat neglected and ontologically austere subcategory of rational motivation theory is what I call "cognitive naturalism." My aim is to examine this type of account, certain of its basic themes and supporting arguments, by considering in some detail one version of it.

1

The basic idea behind cognitive naturalism is that normative and evaluative truth are in some way constituted by the fact that one or more agents would choose to perform an action or respond positively toward a thing were they adequately informed about it. Cognitive naturalist (or CN) theories are "naturalistic" because they make no basic ontological commitments to entities that do not already figure in currently well-confirmed scientific theories or commonsense views about the material world (with reasonable priority being given to the former in cases of genuine ontological conflict). The point here is that ethics has no distinctive or *sui generis* ontology of its own, perhaps, as we shall see, with the limited exception of projections of and from familiar psychological processes that occur every day. Correlatively, such accounts purport to be thoroughly reductivist in just this respect: no concept or expression employed in the analyses provided, nor indeed any complex component in them, is by itself "normative" or "evaluative" in some strong sense. Rather, that certain facts and propositions have these characteristics is a function of the entire combination of elements in the theory's analyses. CN accounts purport to give analyses, or at least "real definitions," of

[1]Following David Zimmerman's terminology in "Meta-Ethics Naturalized," *Canadian Journal of Philosophy* X (1980), pp. 637-662, at p. 652.
[2]Ibid., pp. 651-652.

normative and evaluative notions in terms of truth conditions,[3] and are "realist" in (at least) two senses. First, the truth or falsity of propositions applying these concepts is logically independent of the beliefs of agents regarding those same propositions. Second, normative and evaluative concepts are not guaranteed to hold of an object owing to the result of any determinable and finite verification procedure.[4]

CN theories are also constrained, whatever their details, by two general principles. The first is the Principle of Normative Motivational Neutrality, or NMN:

> No logically consistent desire, aim, or preference can be shown to be normatively unjustified, relative to an agent at a given time, solely on the basis of its causal origin in him or owing to the facts about its object that are independent of the effect on his motivation of his representation and consideration of them.

The second is the Principle of Necessary Practical Relevance of (true) normative and evaluative judgments, or NPR:

> True normative and evaluative judgments which are valid relative to an agent have a practical or motivational relevance *for him* which is noncontingent.

Under NMN, it is at least conceivable that any internally consistent desire or preference might turn out to be rationally justified and hence to give rise to actions that would be, in the strongest sense, normatively justified relative to the agent in question. Under NPR, whatever kind of fact is essentially involved in the truth of a normative or evaluative judgment, that such a fact obtains necessarily implicates the motivational capacities

---

[3]By a "real definition" I simply mean an analysis indicating that feature or fact in reality which most closely corresponds to conditions implicit in the pre-theoretical concept at issue. Such analyses may have greater logical and theoretical detail than, but conform enough in content to, the pre-theoretic notion for us to say that factual and property reference is preserved. Sometimes, of course, they can't be. Thus, the closest thing in reality corresponding to the notion of demonic possession is, I suppose, schizophrenic psychosis, but the correspondence is just not close enough. Instances of the latter do not count as cases of demonic possession. What accounts for this is probably the fact that the ontology of the required causal origin of possession is completely absent in the corresponding psychological notion.

[4]Whether and in what sense there are "objective" values and normative requirements we shall consider later on. See section 5, *infra*.

of agents for whom and relative to whom the content of the judgment is in fact normative or evaluative.

The gambit of the cognitive naturalist is to attempt to account for normative validity in terms of its connection with motivation, provided that he can also specify, reductively, conditions under which motivation is maximally subject to rational qualification and criticism. Given, however, the cognitivist and reductivist requirements, and the constraints of the NMN and NPR principles, the entire programme of cognitive naturalism would seem to be a difficult one to complete.

Consider the following problems that come easily to mind. First, how under cognitive naturalism is genuine incontinence possible? Notwithstanding the considerable efforts of Aristotle and R.M. Hare to explain away the phenonmenon, it seems conceivable, indeed it happens, that one may sincerely and comprehendingly believe that he ought to do one thing and yet intentionally not do it. Second, if the possibility of genuine incontinence is secured on some CN account, how could that same account conform to the NPR principle? If not outright inconsistent, incontinence and the validity of NPR lie together uneasily, to say the least. Third, in general, true claims by an agent that he desires or prefers something are not, without more, normatively justified. How might cognitive naturalists allow for the possibility that actual desires are not self-justifying?

Regarding this last question, it will not do merely to advert to a narrowly decision theoretic model to explicate the ways in which desires and preferences are subject to rational and normative criticism, for decision theory generally takes the intrinsic desires and preferences of an agent as fixed from a normative point of view. Only extrinsic or derivative desires can be criticized, for example, in terms of factual error, sheer ignorance or inferential gaffe.[5] However, many considerations suggest that the normative concepts we now employ allow for the possibility that even an agent's intrinsic desires and preferences can be subject to normative criticism, criticism that is somehow inherently binding upon the agent or otherwise rationally compelling. From those who hold that this is impossible, an argument is needed, and none that is convincing has been forthcoming.

It might appear, moreover, that an adherent of cognitive naturalism would have to produce just such an argument. But while it is true that establishing that intrinsic desires are beyond normative criticism would complete the cognitive naturalist programme in metaethics, doing so is not essential to that programme. It is so far conceivable that all the constraints of cognitive naturalism be met consistently with the possibility of sound normative criticism of intrinsic desires. Indeed, any account on

---

[5]See, e.g., *Impartial Reason*, by Stephen Darwall, Ithaca, Cornell University Press, 1983, Chapter 6.

which such criticism was *not* possible would seem incomplete and unconvincing without an explanation of what in our scheme of normative concepts has led us to think and talk as if it were. Though not a matter of necessity, it would be odd if it turned out that those concepts were rankly incoherent. It is not, however, incumbent on the cognitive naturalist to guarantee that intrinsic desires and preferences are *in fact* normatively or rationally criticizable. It might turn out, for example, that the possibilty of such criticism presupposes the satisfaction of remote or contingent conditions not generally within the ken of those employing normative concepts. And such conditions may not in fact be satisfied.[6] If so, any putative extension of the decision theoretic model to allow for the criticism of intrinsic desires would fail, not the minimal test of internal consistency, but that of global coherence.

2

It will be worthwhile to have a look at a version of cognitive naturalism.[7] Perhaps not surprisingly, my favorite is my own. We may begin by examining two fundamental theoretic definitions around which it is built. These definitions, of the artificial terms, 'good*' and 'ought$_{\omega}$', are not intended to give the meaning of either 'good' or 'ought' in English, but to explicate very special senses, one might say limiting cases, of their ordinary counterparts. 'good*' corresponds to the notion of something's having value in the strongest possible sense; 'ought$_{\omega}$' to the notion of what an agent ought overridingly to do, not, for example, *qua* surgeon, artist, politician, egoist, perfect altruist, or even *qua* moral agent (should it turn out that what one morally ought to do is not of conceptual necessity the same as what one ought overridingly to do).

---

[6]This might happen where, for example, the possibility of such criticism presupposed the existence of cognitive and/or inferential operations which might serve to produce, cancel or modify intrinsic desires, but it turned out that there could be no such operations given the facts about the human nervous system. (I plan to deal with this entire issue in a forthcoming work.)

[7]Although they might not accept all of the constraints in terms of which I have defined cognitive naturalism, several philosophers have put forward theories which seem in large measure to conform to them. Recent works by Richard Brandt, David Falk and Peter Railton are prominent examples. See Brandt, *A Theory of the Good and the Right*, Oxford, Oxford University Press, 1979, especially chapters I-VIII; Falk, *Ought, Reasons, and Morality*, Ithaca, Cornell University Press, 1986, especially the essays, "Fact, Value, and Nonnatural Predication," pp. 99-122, and "Hume on Is and Ought," pp. 123-142; Railton, "Moral Realism," *The Philosophical Review* XCV (1986), pp. 163-207.

Where S is an agent, X a thing, A an action or action type within S's power to perform at time t, we have:

(T) A thing X is (to some extent) good* (relative to S) if, and only if, S would have some positive motivation toward X were S adequately to consider enough facts about X such that his adequate consideration of any further facts about it would not alter the positive character of that motivation.

For 'ought$_\omega$',

(T_ω) S ought$_\omega$ to perform A at t if, and only if, S would choose to perform A at t were S adequately to consider enough facts about that alternative such that his adequate consideration of any further facts about it would not alter his choice.

The foregoing definitions are not by any means adequate as they stand, but will suffice for a moment to give the reader the general idea.[8]

There are several things to note about these definitions. First, there is no requirement that the "facts" in question be either epistemically available to the agent or in some narrow sense "empirical"; all that is required is that they *really be* facts about the thing or alternative in question. Theological facts, though in practice hard to verify, may be acutely relevant to a claim about what one ought$_\omega$ to do. That, for example, an omnipotent Being would punish an act of extramarital sex with eternal torment may quite sensibly discourage impending action. Certainly, it might be argued that there has to be an epistemological dimension here. If a motivationally relevant fact about a thing is, relative to an agent, in principle incapable of being known, that may provide some warrant for not requiring even hypothetically that he consider it. But to sustain this view, one would have to provide compelling reasons for taking that approach rather than the obvious alternative, viz., holding that, although an agent's consideration of any fact, viewed as representable by him in abstraction from epistemic considerations, is required under T and T_ω, some evaluative and normative truths may be, on the CN model, in principle unknowable.

Second, the (projected) process of "adequate consideration" invoked in the definitions must be understood, and be capable of being explicated, reductively, i.e., without essential appeal to any genuine normative or evaluative concept. In relation to T and T_ω, the process principally involves representing facts, making inferences from what one has represented and attending to what one has represented or inferred. In this connection, the following requirements are *stipulated*. Any inferences

[8]The definitions set out in the text are intentionally oversimplified. More technically detailed explications of these notions are included in appendix to this paper.

would alter the category of the relevant agent's motivation (under T) or decision (under T_ω) must be logically valid or, broadly, epistemically sound. If an agent makes an invalid inference such that, had he not made it, his decision or motivation would be different, adequate consideration is not complete. If there is some valid inference, not yet made, such that, were the agent to make it, his decision or motivation would be different, adequate consideration is again not complete. If, finally, there is some additional time an agent might spend attending to some bit of correct information he has, whether acquired by inference or not, such that were he to attend to it for that time, his decision or motivation would alter, the process is also not complete.

Third, the basic difference between 'good*' and 'ought$_\omega$', besides the obvious fact that the latter applies only to actions, is that in the case of 'ought$_\omega$', if it is true that an agent ought$_\omega$ to perform a certain action, then given that he has adequately considered that alternative, he will, must, attempt to act.[9] On the other hand, if it is true that a thing is 'good*' relative to an agent, then given that he has adequately considered that thing, all that is required under T is that he have some positive motivation toward it. Nothing follows immediately about what actions he will undertake. In general, one may want a thing and yet not seek to have or further it. There may be other things one wants even more. We might extrapolate here, although I do not think the inference is by any means easy or simple, and say that what an agent ought$_\omega$ to do is wholly or partially a function of what is good* (or bad*) relative to him.

Fourth, and very important, while we focus here only on two central definitions, the general model they exemplify, facts + consideration yielding motivational response, is not limited to them alone. A metaethics of general application should provide the conceptual resources for analyzing the great variety of normative and evaluative predicates there are. *Prima facie*, this theory does so. By adding different sorts of qualifications and restrictions to T and T_ω, one can indeed begin to account for this variety in a number of ways.

Perhaps the most obvious is by restricting the type of fact an agent is required to consider. For example, one might undertake to explicate the notion of "aesthetic beauty" in part in terms of a restriction indicating which facts about things are, and are not, specifically relevant to the question of their beauty. Thus, an aesthetically beautiful thing may be one that agents would respond positively to upon adequately considering features of its form which can be exhibited in one or more of their sensory

---

[9]Recall that 'ought$_\omega$' applies only to actions within the agent's power to perform at the relevant time.

modalities.[10] In a similar vein, one might undertake to explicate moral claims, for example, in an analysis of the expression 'morally ought', in terms of a restriction on the sort of fact hypothetically to be considered by an agent (or class of agents) to those facts that are suitably "impartial" in that they do not relate to matters of "self-interest" or individual advantage. There are many possibilities here.[11] In both of these cases, the strategy is to account, at least in part, for the particularity of the normative or evaluative concept under analysis in terms of a limited range of facts delimited descriptively.

The range of facts agents are to consider is not the only parameter which may be modified in attempts to provide explications of different normative and evaluative predicates. Another is the intensity of the projected motivational response. Thus, the claim that a painting is "magnificent" or a symphony "great" may involve the requirement that the projected motivational response be a relatively strong one, say, within the class of projected positive responses (always given the adequate consideration of some range of relevant facts) to things of that kind. Alternatively, certain evaluative claims may entail that the projected response, though still broadly "motivational," have in addition a certain phenomenological quality or concomitant. To return to our tentative, and likely defective, analysis of aesthetic beauty, some may feel it *is* defective precisely because it omits any reference to the "aesthetic quality" the response must have, a quality with which one can be acquainted, but which

---

[10]Two problems with the details of this proposal are these. First, there is the need to distinguish aesthetic beauty from, e.g., sexual attractiveness. Second, and more difficult, is the question of how one would handle matters of substance in literary and other "representational" media, e.g., what is said in a novel or the subjects treated in a painting or film. These are not, in themselves, obviously matters of "form," but are clearly relevant to the larger issue of aesthetic merit. Perhaps that is part of the answer, viz., that "aesthetic beauty" is a narrower notion than "aesthetic merit."

[11]Various other qualified normative and evaluative claims can be explicated in a different way, namely, in terms of a condition ingredient in their predicates that relativizes adequate consideration to an antecedently specified desire or interest, actual or hypothetical, or set of them. Relevant facts would again, but more indirectly, be restricted to those that are logically or epistemically relevant to the question of whether the subject of the claim furthers or answers to the posited desires and interests. Claims that might yield to this kind of analysis are, for example, "That is a good white wine.", "M is a good method of torture for extracting information from captives.", "X is a good carving knife.", and so on. (Cf. Paul Ziff's account of `good' in *Semantic Analysis*, Cornell University Press, 1960, Chapter VI.)

cannot be analyzed or understood solely in terms of motivation and parameters intrinsic to it.

While I don't think that there is any such thing as the "aesthetic phenomenological quality," it is open to others who find such an approach palatable to argue for it. The crucial point in regard to cognitive naturalism, however, is that insofar as claims about the beauty of a thing are, as most would hold that they are, normative or evaluative, their truth must somehow be related to the actual or projected motivations of agents. To characterize the relevant response *solely* in phenomenological terms, jettisoning any internal connection between it and resultant motivation, will not work. What could there possibly be in any nonmotivational phenomenological quality that constitutes it as normative or evaluative? Being red has, I suppose, a phenomenological aspect, but not thereby a normative one. To argue in the metaethical case that this quality is *very special*, even though in no respect motivational, rings hollow for several reasons. It again leaves open the clear possibility that normative and evaluative facts are, even for thoughtful well-informed agents fully acquainted with them, no more relevant to our actions and plans than random facts about colors or prime numbers. Moreover, that the phenomenological quality in question is "special," and hence elusive, is implausible. Many who have no trouble applying normative and evaluative concepts are unaware of any such quality. It can't be that it is just hard to perceive, for assuming that I and others have, and are aware that we have, any number of true evaluative beliefs derived in large measure from our experience of things, we must have been, all along, acquainted with this quality day in and day out, but can't locate it on any map of phenomenological qualities of which we are aware.

I might add here that it seems to me a bit late in the day to be raising the question, almost *ab initio*, of whether values can be understood on the model of secondary qualities, e.g., colors. If cognitive naturalism can claim for itself the explanatory advantages I think it can, there will be little need to develop, nor much hope of success in developing, a new kind of metaethics on the secondary quality model. In my view, the best one can say is that there are some similarities between value and secondary qualities, and some profound dissimilarities. For example, concerning the latter, there is no *sui generis* sensation or perceptual quality conceptually linked with correct ascriptions of the corresponding secondary quality, here, goodness. The closest thing available are positive motivational responses which, just in themselves, are very much unlike sensations. Moreover, it would be stretching things quite a bit to say that the analog of "normal conditions" for secondary quality ascriptions here could be the adequate consideration of sufficiently many facts. One could argue for these views, but my point is that the secondary quality "model" will eventually drop out as irrelevant. Aside from vaguely gesturing at the

needed details of a developed cognitive naturalist theory, no advantage is gained.[12]

Another recent metaethical alternative to cognitive naturalism is Simon Blackburn's "quasi-realism" or "projectivism." I can see no advantage there over cognitive naturalism either. Quasi-realism is a sophisticated version of noncognitivism with the added hypothesis of a certain psychological process to explain why we erroneously talk about values as properties that objects external to us have, and therefore derivatively, err in being metaethical realists and cognitivists. Blackburn writes, "On this view [which he attributes to Hume] we have sentiments and other reactions, caused by natural features of things, and we 'gild or stain' the world by describing it as if it contained features answering to these sentiments, in the way that the niceness of an ice-cream answers to the pleasures it gives us."[13] Nevertheless, cognitive naturalism, which, recall, is both cognitivist and realist, if it can explain the intrinsic relevance of normative and evaluative facts to agent motivation, has a general advantage over all forms of noncognitivism, namely, it can subsume their explanatory successes and intuitive appeal without at the same time having tortuously to explain away the fact that we make genuine claims to truth and knowledge regarding normative matters. Moreover, a second major motivation for noncognitivism, that cognitivists are inevitably faced with an intolerable choice between an implausible descriptivism on the one hand, and mysterious nonnatural properties, on the other, has no force against CN accounts, as we shall see.

One of the principal grounds on which Blackburn himself rejects cognitivism and realism in ethics is his view that the notion that normative

---

[12]John McDowell has taken some tentative steps toward developing such a model for normative and evaluative concepts. See his "Values and Secondary Qualities," in *Morality and Objectivity, A Tribute to J.L. Mackie*, edited by Ted Honderich, Routledge & Kegan Paul, 1985, pp. 110-129. Simon Blackburn and Alan Goldman have voiced some rather stiff criticisms of the approach. See "Errors and the Phenomenology of Value," by Simon Blackburn, in Honderich, *op. cit.*, pp. 1-22; and "Red and Right," by Alan H. Goldman, *The Journal of Philosophy* 84 (1987), pp. 349-362. A careful reading of Goldman's paper will indicate that cognitive naturalism is quite immune to the difficulties he raises for McDowell, as it is also immune to related objections advanced by Warren S. Quinn against "moral realism." See Quinn's "Moral and Other Realisms: Some Initial Difficulties," in *Values and Morals*, edited by Alvin I. Goldman and Jaegwon Kim, D. Reidel, 1978, pp. 257-273.

[13]Blackburn, *op. cit.*, p. 5. See also his "Rule-Following and Moral Realism," in *Wittgenstein: To Follow A Rule*, edited by Steven Holtzman and Christopher M. Leich, Routledge & Kegan Paul, 1981, pp. 163-187; and *Spreading the Word*, Oxford University Press, 1984, especially Chapter 6.

and evaluative properties somehow supervene on first-order natural facts about things is incoherent.[14] In my view, Blackburn's chief mistake regarding supervenience is that he assumes that if a thing's having value is a function of certain of its natural features, it is *solely* a function of them. However, such a strong conception of supervenience is not needed to explain the linguistic and conceptual data. Plausible analyses of evaluative supervenience have been developed on analogy to natural dispositional properties. Consider the account by John Campbell and Robert Pargetter:

> Can we use the model we have used to express the relationship between fragility and its natural basis to explicate the nature of the relationship between moral and natural properties? This would come to the following. When we say that some state S is good because it is pleasurable, we mean the following:
>
> (3) being good = having some property which is responsible for being such that <...>.
>
> and
>
> (4) the property which is responsible for S's being such that <...> = being pleasurable.
>
> On this account goodness is a second order property. It is the property of having a property which is responsible for S's being such that it exhibits what we might call "goodness phenomena," and in virtue of which the goodness phenomena are to be explained. So we have goodness, goodness phenomena, and the natural basis for goodness linked by the two identities. And...(3) will be a necessary truth and (4) will be contingent.[15]

On my version of cognitive naturalism, there may be some legitimate doubt about the existence of "goodness phenomena," a description of which would be inserted in '<...>', unless what could count here is the content of definition T itself. There are not going to be any "phenomena" conceptually linked with 'good*' as there may well be in the case of 'fragility' or 'solubility' or even 'digestibility'. That there is in fact nothing else in or about reality that would affect one's positive motivation toward a thing cannot sensibly be understood in terms of a set of characteristic phenomena or occurrences, so the analogy may not be perfect. Even so,

---

[14]See his "Moral Realism," in *Morality and Moral Reasoning*, edited by John Casey, Methuen & Co., 1971, pp. 101-124, especially pp. 105-116. For a reply, not all of which I agree with, see "An Alleged Difficulty Concerning Moral Properties," by James C. Klagge, *Mind* 93 (1984), pp. 370-380.

[15]"Goodness and Fragility," *American Philosophical Quarterly*, 23 (1986), pp. 155-165, at pp. 161-162.

T and T_{ω} otherwise fit the model quite well. In (3) and in (4), the righthand component in definition T would be put into the brackets, i.e., <S would have some positive motivation..., etc.>. That being done, then if the analysis given in T is correct, indeed (3) would be necessary and (4) contingent.

3

As we have seen, it is a fundamental tenet of cognitive naturalism that the specifically normative and evaluative character of predicates, concepts, propositions, etc., can be understood jointly in terms of the projected motivations of agents relative to whom their application is indeed normative or evaluative, *and* the requirement, more or less extensive, that such responses have been produced by or survived the adequate consideration of some range of facts about the action or thing in question. It is vitally important to understand that *both* components are essential. The motivational component serves to account for the "guidesomeness" of normative and evaluative claims. Such claims and, if true, the facts they indicate, are not inert, or at least, are not as contingently related to action as are facts we ordinarily, but perhaps obscurely, think of as being merely "descriptive." On the other hand, as we have seen, normative concepts do not simply duplicate the function of standard motivational concepts, for example, of a desire, goal, choice, preference, decision, and so on. Motivational concepts alone cannot provide a model for the analysis of normative and evaluative concepts. Such an approach could not account either for the possibility of genuine incontinence or the normative criticism of desire. The requirement of adequate consideration as a qualification of motivation is intended to overcome these deficiencies in the simple subjectivist model. But to come full circle, one cannot go so far in allowing for such criticism as to sever all connection between motivation and normative concepts, that is, to take an extreme objectivist or externalist position. In that case, one is again left with the seemingly insurmountable problem of explaining the inherent "normativity" of normative claims and facts.

We can now define the notion of "CN-Normativity":

> An application of predicate P is CN-normative (or evaluative) relative to an agent (or class of agents) S and a thing X if, and only if, the truth of the proposition, P(X), logically implies that S would have some positive or negative motivation toward X upon adequately considering some range R of facts about X.

A principal substantive claim of cognitive naturalism is that CN-normativity is an adequate explication of our pre-theoretic notion of normativity.

Having considered in very general terms a number of ways in which the scope of adequate consideration might be restricted so as to yield analyses of qualified normative concepts, we might return for a moment to definitions T and T_ω and raise the question of why 'good*' and 'ought$_\omega$' occupy a central place in our account. The answer is simple. It is precisely because they are defined without any restriction on the scope of adequate consideration. Potentially, *any fact* about the world is relevant to a claim that a thing is good* or ought$_\omega$ to be done. This has an important consequence: relative to a given agent with respect to whom it is true that a thing is good* or a certain action ought$_\omega$ to be performed, it follows that if he does not have the positive motivation (required under T) or does not at least attempt to perform the action (as required under T_ω), then he is ignorant of some fact about the world, some truth, that would be decisively relevant to him were he to be aware of and sufficiently consider it. Conversely, if he has achieved adequate consideration as required under T and T_ω, then there is no fact about the world, no consideration, no truth, no part of reality, the representation and consideration of which by him would alter, respectively, his positive motivation or decision toward the thing or action in question.

It is now also possible to indicate why it is not an essential thesis of cognitive naturalism that intrinsic desires be immune to rational or normative criticism. This is because it is left a logically open question whether even intrinsic desires might be extinguished or their strength significantly reduced, or entirely new ones generated, in the process of adequate consideration. What psychological mechanisms could account for this I do not know. But if such a thing ever occurs, and there is so far no argument showing that it cannot, facts relevant to normative matters will not of necessity be limited to those relevant to determining which actions will maximize net expected utility *relative to* the entire set of the agent's *actual* intrinsic desires and preferences.

Under cognitive naturalism, the degree to which desires and choices are normatively justified is a function of whether they would survive or be produced by the consideration of motivationally relevant truth, the more extensive the latter, the greater the degree of normative justification involved. No special or *a priori* normative status is given to any descriptively specified subcategory of fact. Nevertheless, within this theory, one has the conceptual resources to claim, as a substantive empirical matter, that facts falling within any descriptively given category would actually be of overriding motivational importance in the context of projected adequate consideration. But in doing this one must pay an epistemological price, a price consistent with any genuine commitment to

realism, even in metaethics. One must have a sensible answer to the question, "How do you know?"

Some may feel that cognitive naturalism is fatally flawed because it is still, in every instance, too relativistic and subjectivist to be faithful to pre-theoretic normative and evaluative concepts. In T and T_ω, and in virtually every suggestion made above concerning possible ways of analyzing qualified normative predications, there appeared, explicitly or implicitly, the rider, "relative to [an agent] S". But clearly, not all normative concepts are *that* relative. Otherwise, one couldn't make much sense of the fact that different people disagree and argue about the truth of the same normative proposition. And even if this basic "no conflicts" objection is avoided once we turn to the more complex definitions, T' and T_ω',[16] there is still no way on this model to formulate normative claims that purport to have wide intersubjective validity. But we do sometimes wish to make such claims. Pending an argument that they all must be incoherent, cognitive naturalism should be rejected.

This would be a troubling argument except for one thing. Nothing is easier under cognitive naturalism than to reconstruct normative claims purporting to have any degree of intersubjective validity one likes. For example, if one wished to provide an analysis of an overriding ought-claim which, if true, is normatively valid for all rational agents whatever, all one need do is add to T_ω the requirement, as a truth condition, that all rational agents, including S, would prefer S's doing A at t to his not doing A at t, upon adequately considering the matter. Using this new predicate, if one agent asserted that S $ought_\omega$ to perform A and another denied this, it is now logically guaranteed that their pragmatically opposite claims are genuinely inconsistent. By adding still further conditions, for example, a restriction on the range of relevant facts that they include none pertaining to mere self-interest or personal advantage, one might have a model for moral claims purporting to have universal intersubjective validity. Finally, if one wished to assert that this latter sort of claim was, in addition, normatively overriding, one need only additionally assert that the positive motivation that would be generated or sustained in every rational agent by his or her adequate consideration of "impartial" facts regarding S's performing A at t would override any contrary (partial) motivations that would be generated in the context of adequate consideration of the sort required under T and T_ω, viz., factually unrestricted adequate consideration.

One may here have the beginnings of an argument that cognitive naturalism, *qua* foundational account in metaethics, can defeat, or better, subsume, a number of its rivals in the category of rational motivation theories. Consider all those accounts, whether contractarian or not, which

---

[16]See the "Appendix", *infra*.

purport to establish that overriding normative, more often than not specifically "moral," requirements can be derived from or discovered by determining those choices we would make from behind a veil of ignorance. Such claims can be reconstructed within the CN framework as follows. The restrictions on information defining the veil in question serve to establish what more specific normative predicate or concept is under analysis, e.g., 'moral', 'is just', etc. Requirements of unanimity of choice establish the degree of intersubjective validity at issue. Finally, the claim that such choices, or the principles they select, are *overriding* would have to be, under my theory, a claim to the effect that the motivations or choices generated by the consideration of information not excluded would remain dominant once the veil is lifted. And indeed, one finds that this last condition is often invoked in such accounts, though its satisfaction is just assumed. For example, Rawls posits an effective "sense of justice" that will sufficiently override inclinations contrary to the requirements of justice. Regarding actual practice, as it were, unveiled, that *may be* only wishful thinking.

That one can, in these and many other ways, reconstruct normative and evaluative propositions having any degree of intersubjective validity again illustrates the versatility of cognitive naturalism. Yet there is, as I have already suggested, an epistemological price for deploying these various semantic resources. To assert warrantably a normative claim purporting to have, for example, universal intersubjective validity, one needs to have good epistemic reasons for thinking that it is indeed *true* that all rational agents would respond similarly to the thing or action in question. Quite obviously, such claims require a much stronger or more extensive evidentiary grounding than normative claims explicitly relativised to a single agent or small group of agents. But what is wrong with such a consequence? Within a realist framework, broader claims naturally demand more extensive justification.

If some, still holding themselves out to be "realists" in metaethical matters, find this unpalatable, our theoretical suspicions surely ought to be aroused. We would then be confronted not only with intuition-based requirements regarding the *content* of normative propositions, often accompanied by a medley of intuitions about which such claims *must* turn out to be true, but now with the additional constraint that these claims can be, indeed are, known easily or intuitively, e.g., "from the armchair." Pressed with enough vigor, this triad of demands can I think be shown to threaten realism in ethics.[17]

---

[17]Nor will it do, I think, to advert here to some such notion as "wide reflective equilibrium" even as a methodological alternative to cognitive naturalism or as a procedure supposed to be neutral between metaethical accounts. As I have argued in detail elsewhere, reflective equilibrium, sensibly interpreted, collapses into cognitive naturalism. See my "Anti-

Also unavailing as an objection to our proposed treatment of intersubjective normative validity is the (from my experience) often repeated claim that ordinary normative and evaluative concepts are in no sense whatever relativistic. The objection then is that cognitive naturalism is inadequate because it allows the use of normative and evaluative expressions in judgments that are highly relativised, even to single agents, once explicit or implicit intersubjectivity conditions are suspended. But our ordinary concepts do not permit that sort of suspension.

The main problem with this objection is that it wildly ignores linguistic reality. One often hears discourses in which normative judgments are explicitly relativised and maintained, usually by one or more parties to an extended dispute about some such matter. For example, a common remark is something like, "Well, from my perspective it is wrong to do that sort of thing, but not from yours. I guess we're just basically different in this." People now readily reach the point, perhaps prematurely, of giving up on the working assumption that the opposing party to a substantive dispute hasn't got his facts straight, or hasn't got all the relevant facts, or about them has reasoned badly or not enough. To respond that this only lamentably shows how the weeds of metaethical relativism have spread through the culture won't suffice, for the issue at hand is the content and character of normative concepts now in use.

Indeed, to point toward an outwardly less relativistic age regarding normative matters, say the Victorian, does not show that the English had fundamentally different normative concepts than we do now. What may better explain the relative absence among them of normative relativizations in speech and argument is, rather, the prevalence then of widely held beliefs which, while not in themselves value judgments, nevertheless had implications for the issue of intersubjective normative validity. To take one instance, the sort of belief that might account for this is the theological one that an omnipotent God gave all men and women the same basic character. Hence, all would respond alike to things if properly educated or informed. (Views very much like this did underlie many late nineteenth century reform movements in Great Britain and the United States.) Another alternative is the specifically metaethical view, then widely inculcated through education and cultural conditioning, that

---

Intuitionism and Reflective Equilibria Revisited," *Pacific Philosophical Quarterly* 69 (1988), pp. 201-221. For a general sketch of wide reflective equilibrium, see "Wide Reflective Equilibrium and Theory Acceptance in Ethics," by Norman Daniels, *The Journal of Philosophy* 76 (1979), pp. 256-282. For an (I think unsuccessful) attack on Brandt's version of cognitive naturalism from this perspective see Daniels' "Two Approaches to Theory Acceptance in Ethics," in *Morality, Reason and Truth, New Essays on the Foundations of Ethics*, edited by David Copp and David Zimmerman, Rowman & Allenheld, 1985, pp. 120-140.

what is valuable or right is not relative to agents, groups or societies. This is not itself a judgment of the rightness or value of anything. Persons of widely different cultures might agree on the foregoing point and yet disagree on every substantive evaluative matter.

That the tides of relativism rise and fall historically lends credence to my view that what accounts for this are not inexplicable shifts from one set of basic normative concepts to another, but rather culturally variable correlative hypotheses on which the likelihood of ultimate disagreement is different. Moreover, the sort of potential *value relativity* embraced under cognitive naturalism is not the same as *value relativism* actually asserted. The latter is the view that *in fact* there are significant areas of unresolvable disagreement over normative and evaluative matters. Cognitive naturalism, as an analytical theory, entails neither this view nor its denial.

It should be noted here that adding to a normative claim conditions that, if satisfied, will result in some measure of intersubjective validity does not guarantee that the normative requirement expressed by the claim has a greater degree of normative validity or force relative to a single agent who falls within the relevant class. Under cognitive naturalism, intersubjective normative validity is only "horizontal." The strength of an agent's positive response toward or preference for a certain course of action upon adequately considering the matter is not of necessity increased should it turn out that all other agents would prefer his taking that course upon their considering the matter. Such a thing might occur. That is, the very fact that others would prefer that an agent perform a certain action might, under adequate consideration, make that alternative even more attractive to him. On the other hand, it might not. Conversely, the mere fact that one or more others would disprefer a given agent's performing an action does not entail that that agent's decision to perform it will disappear upon his adequately considering the matter, including, if relevant, the fact about others' contrary, but informed, preferences. Once we have reached the level of cognitive qualification relative to an agent required under T and T_{ω}, requirements expressed by true propositions which are CN-normative with respect to him are as normative as they can get. Facts about the informed preferences of others, if motivationally relevant *to him*, would have already been factored into the process of adequate consideration.

4

An important objection at this point is that cognitive naturalism allows the possibility that virtually any practice, no matter how horrible, might turn out to be overridingly normatively justified for an agent, and such a consequence cannot be countenanced, even as a mere logical possibility.

How can what is normatively justified or unjustified be held hostage[18] to the potentially horrendous preferences and choices of perhaps one depraved, but adequately informed, agent? Surely normative pluralism has some reasonable limits!

What is being called into question here is the soundness of the NMN principle. Moreover, there are really two thrusts to this objection, and one of them rests on a misunderstanding. An agent or class of agents is not constrained in what they justifiably seek to do or prevent by the adequately considered, but contrary, choices of others. On the CN approach, there is no required commitment to unlimited tolerance or pluralism in regard to any matter. If normative justification, whether overriding or qualified, is potentially *relative to* the projected informed responses of different agents or groups of agents, then it *may* turn out that one agent S would be fully justified in attempting to perform an action A, while at the same time others would be fully justified, not only in choosing that they themselves not perform actions of that type, but in preventing S from doing so as well.

This issue aside, the deeper question raised by the objection remains, namely, how can it be that, even relative to a single agent, certain horrible actions might turn out to be normatively justified? It is necessary, however, to take care to isolate just what it is that is motivating the objection, for some quite legitimate concerns have no bearing here against cognitive naturalism, or in particular, my version of it. Nothing in my account rules out the possibility that our dearest *substantive* normative judgments are true, or that our most altruistic or benevolent inclinations would not only survive, but become dominant in the process of adequate consideration. The case, such as it is, for these claims is not weakened by cognitive naturalism. Unfortunately, their truth is not logically guaranteed either. It is a substantive and in some respects empirical matter whether one agent, or all agents, on adequately considering the matter, would refrain from committing acts of indiscriminate brutality. But why should the provision of such a guarantee constitute a theoretically necessary condition for having a successful metaethics? Consider a parallel objection that might be offered against an analysis of color terms, for example, the term, 'red', that from the analysis alone one cannot deduce that the plastic cup sitting before me as I write is red, or that the heaviest object in the upper left drawer of my dresser is not red. Or again, consider an objection to an account of truth on the ground that it did not guarantee that, say, the General Theory of Relativity is true. Neither of these would constitute a compelling objection to those accounts. Surely our intuitions warrant that there is some sensible distinction to be maintained, even in normative matters, between conceptual and/or metatheoretical claims, on the one hand, and substantive ones, on the other.

---

[18]This echoes a complaint made against Brandt in Daniels, *op. cit.*, 1985.

Accepting the latter claim, of course, does not settle the matter. Why not say, for example, that in some cases, beyond certain generous limits, some kinds of actions, specified descriptively, just cannot, a conceptual 'cannot', count as normatively justified even relative to one or a few agents? This maneuver, directly to build certain descriptive requirements into normative language, faces a dilemma. If such descriptive limits on preferences and actions are not somehow combined with links to agents' motivational capacities as required by the NPR principle, normative facts become implausibly inert, or if being inert is not implausible, then at least the whole question of why one should bother or care about what is "normatively required" is opened up again. Alternatively, if descriptive limits are combined with conditions satisfying the NPR principle, one may thereby gain a logical guarantee that it is not the case that brutal acts are normatively justified, but by no means does one thereby gain a like guarantee that they are normatively *unjustified*. To see this, consider what such a combination of descriptive and motivational conditions might look like. Perhaps this: "Performing an action is normatively justified if and only if *both* [one, some, most, all] agents would choose to perform it upon adequately considering the matter, and the action in question is not of any of the following kinds - indiscriminate brutality, ... etc." What this leaves us with is plain to see. To the extent that it could turn out that brutality is normatively justified given the truth of cognitive naturalism, then to precisely the same extent it might turn out under this sort of hybrid account that it is not unjustified. All that is gained in deploying the hybrid is a potentially larger logical space between the contraries, 'justified' and 'unjustified', than there was before. And within that space, under whatever name it might go, there will remain, certainly for individual agents, the practical question: What is to be done?

An approach some might find more satisfying than the first, perhaps because more subtle, is to invoke normative concepts used directly to assess the character or motivational capacities of agents. For example, suppose an agent S truly asserts of himself that he ought$_{\omega}$ to do A, where A is some act of extreme brutality to be done for its own sake or for fun. Let us also assume that "we" would overridingly prefer that S not perform A upon our adequately considering the matter. In such a case, and given the considerations we have already examined, on what basis could we claim that S's decision to perform A is normatively unjustified, not say, just relative to us?

Could one say, and so be done with the issue, that in deciding to do A on adequately considering the matter, S has thereby shown himself to have a depraved character, and because his decision results in part from such a character, it cannot be normatively justified even relative to him? But what does this thesis involve? It must be something more than a reiteration of the fact, already granted, that our contrary preference regarding S's performing A, or perhaps indeed our taking action to prevent

S from doing so, is justified relative to us. In fact, a proponent of this approach needs to build two bridges to get to any significant conclusion here, one bridge from the ascription of "depravity" to the lack of justification of S's decision *tout a fait*, and assuming that is accomplished, a second from this established lack of normative justification to something in that fact that S *must be* concerned with in the sense that he cannot rationally or coherently ignore it.

Regarding the first task, we are faced again with a familiar dilemma. Either "depravity" is something, in whole or in part, simply descriptively delimited, for example, in terms of a list of decisions or preferences, informed or not, that are simply by definition depraved or show a depraved character, or, the second alternative, the normative significance of depravity is a function of something else. If the former, we are faced with precisely the same epistemic and other problems, now at one definitional remove, considered earlier in connection with the attempt directly to build descriptive inclusions and exclusions on action into the concept of normative justification. If, on the other hand, the normative significance of depravity is a function of something else, that something else has yet to be spelled out.[19] Certainly, pointing out that it hasn't does not constitute proof that it can't, but one's theoretic patience may begin to wear thin, especially where, as here, there is an attractive alternative view available. As in most areas of inquiry, good theories do not come complete with proofs that no better theory is possible. And most important, any succesful competitor to cognitive naturalism must spell out in detail what advantage, explanatory or otherwise, is gained over the CN approach, specifically, in connection with the latter's allowing the *logical possibility* that contrary preferences about the same thing may be normatively justified relative to different agents and leaving it as a contingent fact, if fact at all, that this ever happens.

At this point, what options are left? There is one which might be used not only to solve the problems delineated in the last paragraph, but also to construct the second (and really more important) bridge from the concept of normative justification to the inherent bindingness of valid overriding normative requirements, such that, despite the fact that it is true that S ought$_\omega$ to perform a brutal act A, it nevertheless remains the case that S's decision is normatively unjustified, relative to him or otherwise, and this in some respect which accounts for the inescapable relevance of the latter claim *to him*, whether or not that relevance is somehow manifested or reflected in his informed motivations. The alternative in question is to show that in performing A, S could still be counted as *irrational* no matter how informed is his (positive) motivation regarding A.

---

[19]Of course, it cannot consist of cognitive and inferential qualifications of motivation of the sort involved in T and T$_\omega$, for that would lead us back to cognitive naturalism which we have assumed was to be rejected.

A substantiated charge of irrationality against even an adequately informed S would have profound advantages here. While S's motivation could not be altered by facts and arguments we might present to him, and hence no claim of inescapable relevance *for him* could be made out in terms of any obvious link with his own motivations, we might here have a suitable substitute, viz., that S is somehow being incoherent in choosing and pursuing the action A. This may sound initially promising as a way around or through the agent-centered limits imposed under cognitive naturalism, but there is a danger as well. There is nothing magical about the word 'irrational'. The basic problem faced by the opponent of cognitive naturalism would not be solved should it turn out, as with the possibilities we considered regarding 'unjustified' and 'depraved', that S's choice of A was ruled out solely or principally owing to descriptive inclusions or exclusions built into the concept of rationality, and nothing more. For then, certain aspects of rationality could be coherently ignored by S. Rather, what is needed is a showing that S, in performing A knowing or believing it to be irrational, is thereby being incoherent. But to date, no account has come close to establishing any such thing.[20]

Our assessment of cognitive naturalism ought correspondingly to be enhanced could one establish some such result on its basis. In fact, I think this can be done. Specifically, I think one can show that an agent who (correctly) grants that he ought$_{\omega}$ to do A, but, because he has not

---

[20]Two of the recent, and to my mind better, attempts to show that "moral" requirements, understood as normative requirements that might be contrary to choices one would make in seeking to maximize individual expected utility, even under conditions of adequate or full information, are requirements of practical rationality are Darwall's *Impartial Reason, op. cit.*, and *Morals By Agreement*, by David Gauthier, Oxford University Press, 1986. Concerning the success of either, however, there is plenty of reason to doubt. On Gauthier, see "Morality and the Theory of Rational Choice," by Jody S. Kraus and Jules L. Coleman, *Ethics* 97 (1987), pp. 715-749; on Darwall, see my "Impartiality and Practical Reason," *Philosophy Research Archives* XII (1987), pp. 1-65. I argue that Darwall's project fails, *inter alia*, because at a crucial juncture he simply assumes, contrary to the NMN principle, that it is analytically a feature of the ideally or maximally rational agent that such an agent would have an overriding preference for rationality and rational action no matter how that might affect his other interests and in whomever such rationality might manifest itself. Moreover, even if this were true, and it is not, that may simply show that "rationality" is too thick or substantive a notion to do the required job, namely, establishing that an agent who is under a "rational," or whatever, fundamental normative requirement, is under a requirement that is inescapable in this sense, that if he acknowledges but fails to act in accordance with it, he can be shown to be incoherent.

completed the process of adequate consideration, has contrary intentions, thereby commits himself to an incoherence in his beliefs, desires and intentions. The key to showing this would involve demonstrating that, under the relevant hypothetical conditions, the object of his intention in fact cannot relevantly be what he takes it to be, where what he takes it to be is decisively his reason for and the subjective ground of his pursuing it.

Briefly to sketch this argument, I shall first assume that an agent S ought_{ω} not to perform a certain action A. Let us also suppose that S firmly believes this and yet intends to perform A. It so far follows that S has not achieved adequate consideration of the alternative A. But in these circumstances, where is the incoherence in what S is doing? There is some "object" of S's intention to perform A. For example, S may have an intrinsic preference to perform A, or may believe that performing A is necessary (or the least costly way) to bring about some state of affairs which is intrinsically desired or preferred by him. Here the content of the relevant intrinsic desire or preference (or set of them) indicates the object of S's intention. What is vital to note is that S's intention is not based on *any* desire or motivation he has that is overriding in this sense: he would have and/or act on it *no matter what* further facts about the world he might acquire or consider. This would be contrary to our assumption that he ought_{ω} not to perform A. But it follows then that the object, O, of his intention to do A, under the descriptions of O he now accepts, is not what he relevantly takes it to be, *as he must admit*. How so?

There are only two possibilities. First, some of the motivationally relevant beliefs he has regarding O are false, and hence, O is not, in a relatively straightforward sense, what he takes it to be. Because S is logically committed to granting this, he must also grant that that he is pursuing O in or by performing A because O is, say, F, and yet it is not. This is incoherent. Second, it may be that while all the beliefs that S so far has about O are true, there are other facts about it, and hence about his performing A, which would alter his intention to perform A *and* which are such that no further facts would change it back. But in that case, and given that O is not (and cannot by hypothesis be) intrinsically and overridingly desired by S in the sense already indicated, what is the object O of S's current intention to perform A? It is O, under the true descriptions S now accepts, together with the following condition: "O, *provided that* there is nothing else about O that would make me [S] prefer not to perform A." But since, on our assumptions, facts exist which guarantee that this last condition is not satisified, it follows that, as S knows, S would prefer to act only if the open condition is satisfied, and yet must grant that it is not, and still intends to perform A. That is again incoherent.

But why is it essential to the coherence of his intention and impending action that S not know that the open condition is not satisfied? That is the crucial question here. Well, from what has already been said it is simply a fact that S's intention is conditional with respect to the state of the world

beyond the specific true beliefs he now has regarding O and A. This *is* the fact expressed by the open condition. On all the assumptions we have made so far, S is in the position of having to say or admit just this: "I desire O because it is F (or in the limit, just because it is "O" or "A"), but not come what facts there may. But I grant that it is guaranteed that facts of the relevant sort obtain." In essence, S wants an O which is not relevantly the O he wants. The second possibility therefore also leads to a practical incoherence.

If the foregoing argument works, or can be made to work by suitable modification and expansion, we have quite an interesting result. Failure to do what one knows or grants one ought$_{\omega}$ to do is irrational in the most fundamental sense: it is incoherent. This holds for any S and A, and holds without positing any substantive desire in S whatsoever, e.g., a desire to be "rational." Moreover, only 'ought$_{\omega}$' guarantees this result (with the possible exception of a normative predicate that unnecessarily requires full omniscience of the relevant agent). We may now begin to understand why it is a genuine and open question whether "moral," or indeed any other 'ought's are rationally overriding.

5

A final matter we shall address is the general objection to cognitive naturalism that normative and evaluative concepts are in content more objective than it allows. As we shall see, it is sometimes hard to get a handle on what is involved in the view that they are "objective" in some respect that defeats at least the reportive success of my account.

Consider John Mackie's widely known discussion and rejection of "objective values" in Chapter 1 of *Ethics, Inventing Right and Wrong*.[21] It is most instructive to note how Mackie himself is pulled in different directions in his attempts to explicate normative objectivity. In fact, in a number of places he so characterizes the notion that definitions T and T$_{\omega}$ would suffice. For example, he writes: "So far as ethics is concerned, my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid." (29) Presumably, if there were some such "categorically imperative element," there would be an objective value. But what sort of element is in question? "A categorical imperative, then, would express a reason for acting which was unconditional in the sense of not being contingent upon any present desire of the agent [or perhaps *any* agent] to whose satisfaction the recommended action would contribute as a means - or more directly: 'You

[21]Penguin Books, 1977. Parenthetical references in this section of the text are to pages in this work.

ought to dance', if the implied reason is just that you want to dance or like dancing, is still a hypothetical imperative." (29) Just so, that an agent ought$_\omega$ to perform an action is likewise not contingent in these respects. The truth of such a claim does not presuppose or posit any present desire in any agent, including the subject of the claim.

Referring to any arguments that support an evaluative conclusion, Mackie puts his central point another way: "...what I am saying is that somewhere in the input to this argument - perhaps in one or more of the premisses, perhaps in some part of the form of the argument - there will be something which cannot be objectively validated - some premiss which is not capable of being simply true, or some form of argument which is not valid as a matter of general logic, whose authority or cogency is not objective, but is constituted by our choosing or deciding to think in a certain way." (30) Yet, as before, the truth of a claim that one ought$_\omega$ to perform a certain action does not depend on our choosing or deciding to think in a certain way, at least in any relevant sense of 'think in a certain way.' Of course, the truth of a sentence token of the type, "This rock weighs more than 2 kilograms.", presupposes in some way that "we" have and apply the concepts invoked by terms in the claim, e.g., of a rock, a kilogram, what it is to have a certain weight, and so on. But *that* can't be what makes a claim not "objective."

Finally, Mackie claims that the ordinary person uses moral language to characterize an action as it is "in itself" and not to make claims about or to express his, or anyone else's, attitude or relation to it. "But the something he wants to say is not purely descriptive, certainly not inert, but something that involves a call for action or for the refraining from action, and one that is absolute, not contingent upon any desire or preference or policy or choice, his own or anyone else's." (33) The claim that someone ought$_\omega$ to perform an action is not about that action as it is "in itself," but neither is it simply about or expressive of anyone's attitudes. The truth of such a claim does, however, have something to do with certain relations holding between some one or more agents and the action in question. But again, its truth is not contingent upon anyone's *actual* desires, preferences, etc.

Some may think that the proper rejoinder to my response to Mackie is to say that he should not have limited to the *present* desires of agents those "contingent" factors which rob a normative or evaluative claim of its status as objective. There is in fact some textual support for the view that he did not embrace any such limit. In discussing what the ordinary man is concerned with when considering the objective morality of doing research on bacteriological warfare, Mackies states, "The question is not, for example, whether he really wants to do this work, whether it will satisfy or dissatisfy him, whether he will *in the long run have a pro-attitude towards*

it,..." (33)[22] The last-quoted clause implies that the "pro-attitude" in question need not be a present one. Later on, again attempting to mark off genuine from ersatz objective values, Mackie muses, "How much simpler and more comprehensible the situation would be if we could replace the moral quality with some sort of subjective response which could be causally related to the detection of the natural features on which the supposed quality is said to be consequential." (41) With suitable restrictions on the sort of causal relation at issue, e.g., that the relevant subjective response be motivational and occur as a result of the process of adequate consideration, this is close to a description of 'good*' and 'ought$_{\omega}$'.

I want to claim, nevertheless, that either Mackie's conception of objective value is itself *internally* incoherent, and hence implausible as a reportive analysis of evaluative objectivity, or, that the sort of facts involved in correct applications of 'good*' or 'ought$_{\omega}$' could, if universally intersubjectively qualified, count as objective values. Mackie would likely disagree with both disjuncts. With the first because he seemed to think that, while there is nothing in reality answering to our concept of objective value [his "error" theory], at least the concept itself is coherent even if what it is of is a bit strange. (40) Probably with the second because intersubjective agreement in what is valued, even if universal, is not on his view sufficient for the objectivity of values, though under the right conditions, it might well be necesssary. This strongly suggests that claims that might be constructed from the definitions of 'good*' or 'ought$_{\omega}$' by adding truth conditions requiring convergent responses across agents cannot count, even if true, as indicating something about objective values.

To begin the argument, it is important to bear in mind that Mackie rejects descriptive naturalism as a possible account of objective values. (23) Normative and evaluative terms, insofar as they are used to refer to genuinely objective values, are not merely "descriptive of" natural features of actions and things. (32) Accounts of the latter sort fail to capture the full content of the notion of objective value: "On a naturalist analysis, moral judgments can be practical, but their practicality is wholly relative to desires or possible satisfactions of the person or persons whose actions are to be guided; but moral judgments seem to say more than this. This view leaves out the categorical quality of moral requirements." (33) Claims about what is objectively valuable, including moral claims, involve in part a "...claim to objective, intrinsic, prescriptivity..." (35), a "categorically imperative aspect...". (33) But what is that?

Surprisingly, an answer to this last and rather crucial question is not spelled out in any detail by Mackie. Indeed, his entire direct elaboration of "categorical prescriptivity" is made in two brief characterizations of Plato's "Form of the Good." He says that just knowing or "seeing" it "...will

---

[22]My italics.

not merely tell men what to do but will ensure that they do it, overruling any contrary inclinations." (23) Again,

> Plato's Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it. An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. (40)

Now, it seems to me that the very last explication in this passage really is none, and can be forgotten here, for "having to-be-pursuedness built in" adds nothing helpful to the one with which we started, "categorical objective prescriptivity." But in that case, and setting aside the matter of intersubjective agreement in motive or inclination, a possible aspect of objective value which can be reconstructed within the cognitive naturalist framework, what are we left with which sets objective values apart, *qua* thing or fact, from partial or overriding motivations produced in the process of adequate consideration?

One thing it cannot be, given Mackie's own formulations, is that an objective value, or what is involved in a thing's having such value, must be specified wholly independently of any facts about agents, in particular, any facts about their motivational *capacities*, if not their antecedent actual motivations. If categorical prescriptivity is, as Mackie wants to claim, part of the very concept of objective value, then clearly in the explication of that concept, or implicit in its application to things and actions, some reference to agent motivation must be involved. If a thing has objective value, there must be something about the agent or agents in question in virtue of which their knowledge of or acquaintance with that value, or things that have it, generates the required motivation. If the concept both rules in and rules out just this sort of relation, it is incoherent.

How might one avoid this incoherence and at the same time escape the conclusion that the sort of fact indicated by correct applications of 'good*' or 'ought$_\omega$', fully intersubjectively qualified, can count as objective values? While either fact requires something *vis a vis* the motivational capacities of one or more agents, no antecedent desires are posited. Indeed, since on Mackie's own account, some connection between objective value and motivation is necessary, what better than cognitive naturalism explains why it is that the causal power of objective value works its motivational effect by way of an agent's *knowledge of* or *acquaintance with* objectively valuable things? It might not be that way. For example, one can imagine objects that emit strange rays which restructure an

agent's nervous system so that, when he came within a certain distance, he is caused to have certain desires, all this bypassing any cognitive operations in him and deactivating, also without the aid of the latter, any contrary desires. Would the existence of such an object somewhere in the universe thereby guarantee that there was one objectively valuable thing? Not obviously.[23] Moreover, on such a view, one would have no explanation whatever of the supervenience of value on natural facts. Under cognitive naturalism, one has a relatively straightforward explanation of this and of the phenomenon of reason-giving.

Somewhat surprisingly, it seems that cognitive naturalism provides the best available explanation *both* of the logical independence from one's actual desires of a thing's having objective value and the fact that the motivating effect of objectively valuable things proceeds via the cognition of facts about them. T and T_ω may indeed be faithful to our pre-theoretic normative and evaluative concepts even if certain of the latter are "objective" in the ways Mackie thought. In fact, perhaps the only discrepancy between 'good*' and the Platonic notion as Mackie has characterized it is that the latter seems, implausibly, to rule out the possibility of genuine incontinence, while cognitive naturalism, on the other hand, does not.

An additional point I have been postponing may profitably be introduced here. I have said that my version of cognitive naturalism allows for the possibility of genuine incontinence. It does, not only in the respect that an agent might firmly judge or believe that he overridingly ought to do one thing and yet deliberately not do it, but also that he might know this and yet fail to act accordingly. To see this, consider an agent S who overridingly desires and intends to perform an action A. But also suppose that there is a great and near omniscient being G who, as S firmly and reasonably believes, understands both the world and S's capacity to be motivated by the consideration of fact so that G can determine perfectly what S $ought_\omega$ to do. Suppose that G does this and correctly informs S that S $ought_\omega$ not to perform A. On that basis alone, S might well come to believe that this is so. Moreover, there is no obvious incoherence in holding that S, in these circumstances, might know what we have supposed

---

[23]The point is that the fact that an object would have this odd effect does not show, intuitively or otherwise, that it is objectively valuable then and there. One complication, however, arises owing to the NMN principle. Under it, no projected or actual desire that underlies the ascription of *-value to a thing can be ruled out merely owing to its causal etiology in the agent. Hence, if an agent is so affected by an object, quite independently of cognitive and inferential operations, that he now desires it and would continue to do so upon adequately considering it, then *from that time on* the thing in question is, to some extent, good* relative to him.

him truly to believe. Does it follow on definition T_ω (or T_ω') that S in fact will choose not to perform A? No. It is an open question whether this additional information acquired by S, once it is adequately considered by him, will motivate him not to perform A. All that the truth of the claim that S ought$_\omega$ not to do A relevantly requires here is that there exist facts, representable by S, such that were he to consider them, he would choose not to do A, and further, there are no further facts about doing or not doing A that would, upon being adequately considered, change his mind again. There is no a priori guarantee that his knowing all this will actually motivate S not to perform A. It might or it might not.

Even so, it may be possible here, without stretching things too much, to rehabilitate the Socratic/Platonic denial of genuine incontinence alluded to and ultimately accepted by Aristotle.[24] If we make a distinction between judging or knowing that one ought$_\omega$ to do a thing, and knowing what makes it overridingly required, then indeed, under T and T_ω, genuine incontinence is possible only regarding the former, not the latter. Socrates, or Plato, might well reply that genuine knowledge that a thing is required must be by way of full knowledge of the properties of or facts about it in virtue of which it is required. Second-hand knowledge of the sort involved in our example won't do. Although this invokes a stronger sense of 'know' than standardly at work in ordinary English, there is no deep inconsistency here.

Certainly, there are other forms of metaethical objectivism besides Mackie's. For example, one might embrace an extreme objectivism on which normative and evaluative facts do not involve anything like "categorical prescriptivity." On this view, that a thing is objectively valuable or required need have nothing to do with the prospective effect of acquaintance with it on the motivations of agents. Whether agents would choose to pursue it or even respond positively to it is here entirely a contingent matter. Hence, extreme objectivism would have no trouble allowing for the possibility of incontinence. Indeed, it seems that such a theory would have too little trouble here. Except for a fortuitous accident, cases of full blown incontinence should be as common and unsurprising as cases of continence.

This last matter aside, however, we do seem to be running out of options. I can think of only two. First, extreme objectivism might take the form of descriptive naturalism, viz., *identifying* value and normative requirement with complexes of natural features of things, most likely those features which are standardly cited as counting in favor of a claim that a thing is good or required. The problems with descriptive naturalism have accumulated at least since Moore and need not be recounted here in

---

[24]*Nichomachean Ethics*, Book VII, Chapter 2, respectively at 1145b 26-30 and 1147b 15-16).

detail. To my mind, the most devastating of these is that normative and evaluative properties are inexplicably both redundant *and*, following Mackie, inert.[25]

The other alternative for extreme objectivism is to embrace metaethical nonnaturalism. Objective value is now to be identified with a "nonnatural," or at least nonempirical, feature of things which, nevertheless, must also remain inert in the sense that there is no conceptually necessary connection between a thing's having it and anything about agent motivation. Unfortunately, this approach has less to recommend it than did descriptive naturalism. Not only do normative and evaluative features remain problematically inert, as on the latter view, but now two new difficulties arise, as is also well known. One is that the (at least partial) supervenience of value on natural fact becomes inexplicable. The second is that the entire system of normative and evaluative concepts is open, wide open, to Mackie's error theory, for there is not the slightest reason, so far, to believe that the required sort of nonnatural property exists.

6

In conclusion, it seems to me that the primary theoretical motivations for holding an extreme, or indeed any, objectivist position is first, to ensure that true normative claims, or the most fundamental of them, can be universally intersubjectively valid; and second, to maintain a robust distinction between what is desired (or what in fact motivates) and what is justified (or justificatory). Cognitive naturalism answers to both. Thus, as long as there are compelling reasons to remain metaethical cognitivists and realists, we ought to be cognitive naturalists, or at a minimum, need to take a long and sympathetic look at this type of theory.

Appendix

Here I shall set out more complete explications of the central notions involved in my version of cognitive naturalism. For 'good*', we have:

> (T') A thing X is (to some extent) good* if, and only if,adequate consideration of a (minimal) evaluatively complete set of

---

[25]To say nothing of the striking linguistic implausibility involved in certain versions of this sort of account, viz., that what, for example, `good' means predicatively changes as the term is applied to different kinds of things. What descriptively makes a thing a good carving knife is not at all the same as what makes a thing a good melon. See e.g., Paul Ziff, *op. cit.*, pp. 202-203.

> conditions holding with respect to X would generate, or not extinguish if already operative, some positive motivation toward X in those subjects who would have the same kind of motivation (positive, negative or indifferent) toward X upon adequately considering the elements in such a set.

I here add the further condition that 'good*'is properly predicated of an object X by a speaker only if it is presupposed by him that he would respond positively to X upon adequately considering the matter. To assert that a thing *is* good*, without further explicit or implicit qualification being made or operative, logically implies that the speaker would respond positively toward it if adequately informed. Hence, a sentence token of the form, "X is good*", used assertively, is true only if the speaker would so respond. In effect, then, 'X is good*' is (roughly) equivalent to 'X is such that I [the speaker], and all those who would respond as I would upon adequately considering X, would respond positively to X'. To suspend this condition, one need only introduce some qualifying expression, for example as in, 'Relative to agent or agents S, X is good*'. In asserting a proposition which might be expressed by a sentence of this latter form, the speaker is not asserting something which, even if true, would be "CN-normative" or "CN-evaluative" for him. (For a definition of this last notion, see section 3 *supra*.)

For 'ought$_\omega$', we have:

> (T_ω') S ought$_\omega$ to perform A if, and only if, adequate consideration of a (minimal) evaluatively complete set of conditions holding with respect to the alternative A (i) would generate a decision (choice, dominant desire) in S to perform A, or would not extinguish one already operative, *and* (ii) would generate, or not extinguish, a preference that S perform A in those subjects who would have the same overriding preference as S (positive, negative or none) regarding S's performing A upon adequately considering the elements in such a set.

In a case where 'S' is replaced by 'I', the condition, similar to that added in T', that the speaker would choose to perform A upon adequately considering the matter, is made explicit in clause (i). Clause (ii) concerns uses of 'ought$_\omega$' to make or entertain claims about what others ought overridingly to do where such claims, by their content, are also, if true, normative for or relative to the speaker himself. Thus, the truth and/or felicity of an unqualified assertion by S1 that another, S2, ought$_\omega$ to perform a certain action A, presupposes that both agents would prefer/choose that S2 perform A upon their adequately considering the matter. Where S1 knows or believes that this does not hold of S2, he would

not felicitously express his own normative judgment by using 'ought$_\omega$' in a third-person judgment about S2. Rather, this he might do using a token of the form, "It would be good* [perhaps best*] were S2 not to perform A".

In T' and T$_\omega$', and by implication, T and T$_\omega$, the following subdefinitions (and added explications) also hold:

(a) By a "condition holding with respect to" a thing or action X, I mean anything which can be represented as a condition so holding.

There is no restriction here to "properties," either intrinsic or relational. Literally any fact that might be expressed in a 'that'-clause, with or without perceptual exemplars incorporated by reference, may constitute a condition in this sense, for example, 'that X exists in a universe in which grass is sometimes green'. Further, the phrase, 'can be represented', is to be understood independently of epistemic considerations. A fact or condition C is representable, in the limit, by an agent S if it is logically possible that there obtain further conditions such that (1) their obtaining neither logically implies nor postulates any changes in the motivations or motivational capacities of S that would not arise contingently owing to the process of the adequate consideration of fact, and (2) subjunctive conditionals of the form, "Were S to represent C, then...", would not fail to have a truth value in cases where its antecedent does not obtain, i.e., where the conditional is a counterfactual, *because* the antecedent does not obtain.

(b) A "(minimal) evaluatively complete set" of conditions holding with respect to a thing or action X is a (possibly empty) set L of conditions, $C_1...C_n$, such that
(1) X satisfies the conditions in L;
(2) X satisfies no further condition d such that adequate consideration of d together with the elements of L would alter the category of motivation (*re* T') or the decision or preference (*re* T$_\omega$') that would be caused by or remain after adequate consideration of the elements of L alone; *and*
(3) no proper subset of L satisfies clause (2).

Implicitly, sets of conditions that are "evaluatively complete" may differ relative to different agents or even the same agent at different times. With respect to a single agent, there may be more than one such set holding with respect to the same thing at the same time.

Clause (3) in (b) explicates the qualifier, 'minimal', in the phrase 'minimal evaluatively complete set'. Under (3), every element in the set is necessary for that set's being evaluatively complete, i.e., for its satisfying clauses (1) and especially (2) of (b). Without (3), evaluatively complete sets might, indeed all would, contain all manner of motivationally and/or

normatively irrelevant facts with no finite upper bound. Clause (3) eliminates such irrelevancies and is therefore useful in defining the notion, under T' and T_ω', of a justificatory reason, viz., a good reason (relative to an agent or class of agents) for doing or being positively inclined toward a thing is any fact expressed by any element in one or more minimal evaluatively complete sets holding, on the one hand, with respect to that agent or class, and on the other, the thing in question. On this showing, a good/sound justificatory reason is any fact that could form an element in a rationally conclusive case for being positively motivated toward or actually doing a certain thing. In the limit, as also parenthetically indicated in (b), such sets may be empty. This could occur where an agent already has some positive intrinsic motivation or preference for a thing at the time 'good*' or 'ought$_\omega$' is applied to it. If there is nothing about the thing the adequate consideration of which by him would alter this, then the null set is a minimal evaluatively complete set with respect to the thing and agent in question.

Two further points about T and T_ω' should be made. First, a number of the complexities involved are needed to avoid a standard objection to simple-minded subjectivist accounts of value on which, for example, 'X is good' is equivalent to 'I [the speaker] like X'. On this showing, if S1 claims that X is good, while S2 asserts that it is not the case that X is good, they cannot be contradicting each other in the sense that they are advancing logically inconsistent claims. On the present theory, it is possible, though not logically guaranteed, that if S1 asserts a token of the form, "X is good*", and S2 one of the form, "It is not the case that X is good*", the propositions expressed by these tokens are indeed logically inconsistent. This would be the case if in fact it is true that S1 and S2 would have the same (type of) motivation toward X upon their both adequately considering it.

Finally, while there is not space here to argue or explain this in detail, I think that T' and T_ω' embody a number of advantages over their cognitive naturalist competitors. For example, both Falk and Railton, *op. cit.*, leave out much needed detail in their accounts, respectively, of "merit" and "an agent's good." Moreover, both require that an agent (actually or hypothetically) achieve near omniscience. Not only is this unnecessary, for example, in that it suggests that all facts are relevant to every normative or evaluative question, but is theoretically problematic as well. Subjunctive conditionals with impossible antecedents may well lack truth values. But the counterfactual supposition, "Were S to consider everything, then...", is just such an antecedent, no matter what claim about S's motivations appears in the consequent. Under T' and T_ω', however, omniscience is not required in every or in any case, and minimal evaluatively complete sets clearly can have finitely many, indeed none or few, elements. Hence, my account does not require that the infinitary aspects of things be duplicated in cognition.

Brandt's account of "rational desire" and "fully rational action," in Brandt, *op. cit.*, while fairly detailed, builds into these notions too much epistemic baggage, limiting what an agent is (hypothetically) required to consider to all relevant information epistemically warranted or "available" in the agent's society. Two glaring problems with this are, first, that one (or we) could sensibly undertake to increase the odds of doing what is "fully rational," the strongest normative concept in Brandt's arsenal, by stopping the advance of inquiry across the board, thus stabilizing and limiting "available" information so that one has a better chance of taking its measure in the process of "cognitive psychotherapy." Second, available information not only can include, for Brandt, inaccurate and false "information," but such information about a thing might be the only motivationally relevant information that is available! Thus, on his view, what one ought to do may be entirely a function of beliefs that do not reflect reality.

What is motivating Brandt here is likely the idea, certainly correct, that it can sometimes be epistemically rational to accept propositions which are in fact false (but never under that description!), and hence that it can be rational to act on them and so perform an action even though one is in fact laboring under misconceptions about the alternatives in question. That there is such a thing as rationality even though one has false or inadequate information is not at issue here. That we have need for such a notion, or even variants of it - for why is what it is in that sense rational to do always necessarily a function of *all* available information in society? - is clear. But there are stronger notions, those expressed by T' and T_{ω}', on which it is not a tautology that the point of attempting to be fully rational in normative matters is to maximize one's chances of doing what one ought overridingly to do.