CAN ANTIMORALISM AVOID MORALIZING?
In *Dual-Process Theories in Moral Psychology. Interdisciplinary Approaches to Theoretical, Empirical and Practical Considerations,* ed. Cordula Brand. Wiesbaden: Springer VS., pp. 367-385.

Ronald de Sousa
University of Toronto

ABSTRACT. *To **moralize** is to claim to be entitled to impose normative moral standards on persons who either have not already endorsed them, or having endorsed them, fail to meet them. By **antimoralism**, I mean to refer to a view which can roughly be described by analogy with the small government libertarian view in politics. That view holds that government should be kept out of as many domains of private life and decision as is compatible with peace, the protection of individual rights and social harmony. Antimoralism holds that we should reject the hegemony of morality; contrary to what is assumed by most moral philosophers, most questions in life do not require us to rank moral considerations above all others when we make decisions. On this view, the legitimate sphere of morality ought to be strictly constrained. The word 'ought', in the last sentence, signals a potential incoherence: for is this not itself a normative statement, which although it belongs strictly speaking to metaethics rather than morality, could easily be charged with the sin of moralizing? Who am I to say, and by what authority, what **should** and what **should not** be regarded as included in the legitimate sphere of morality, regarded as the domain of what should and what should not be? The argument I put forward begins by sketching the bankruptcy of all existing attempts to find a foundation of morality. In the light of that failure, an approach commends itself which goes back to Hume's "sceptical solution" to the general problem of justifying rational inference and to Nietzsche's injunction to look not to the justification of morality but to its genealogy. Recent work in psychology and neuroscience suggests that emotions are constitutive of moral judgment. This work further implies that morality is not all of a piece, but arises from clusters of emotional dispositions that order themselves into five or six relatively distinct domains of morality. Two or three of these – the avoidance of harm, justice and liberty – are part of the liberal conception of ethics, but this conception excludes three other domains: purity, community-loyalty and hierarchy-authority. Those three are central to the morality of traditional societies and conservative subcultures. The liberal ideology places the imperatives of those domains outside the boundaries of acceptable contemporary morality. To make that argument, however, seems to expose one to the charge that one is guilty of just the sort of moralizing one deplores in traditional systems of morality.*

CAN ANTIMORALISM AVOID MORALIZING?

Reflections on the relation of science to ethics.

Ronald de Sousa

University of Toronto

http://www.chass.utoronto.ca/~sousa

On the assumption that values do not simply supervene on facts, the justification of morality involves an antinomy: On the one hand, *facts* are all that there is; so if normative statements are not justified by facts, then they are not justified at all. On the other hand, arguments from nature, as exemplified, for example, by the "Natural Law" tradition, are bankrupt in the light of evolution by natural selection, the possibility of which rests on natural diversity. It would seem to follow that no philosophically satisfactory justification of ethics is possible at all. Can science step into the vacuum thus created to set up a naturalistic account of ethics?

Several contemporary thinkers, notably Pat Churchland (2012) and Sam Harris (2011), have argued that science, drawing on evolutionary theory, psychology or neuroscience, is now in a position to take over the determination of moral truth. Whether that claim is valid, however, remains a philosophical question; and the claim has elicited vigorous resistance.

The most entrenched arguments against the very project of looking to science to unseat moral philosophy are likely to remind us of the obstacle to all attempts at naturalizing ethics represented by the "Naturalistic Fallacy". That is where I begin.

1. The Naturalistic Fallacy

Among the standard topics most likely to appear on the curriculum of introductory courses in philosophy is the problem of the *is-ought distinction,* or *fact-value gap*. Any attempt to bridge this "gap" is commonly disparaged as committing the "naturalistic fallacy". The term was first used by G.E. Moore in his 1903 *Principia Ethica* (Moore 1903), and has led to a considerable amount of confusion. Moore seemed to be advancing three separate theses. One was closely related to the point made by David Hume that no matter of fact by itself entails an 'ought' statement. The second thesis was that evaluative terms such as 'good' are indefinable. The third thesis was that evaluative properties such as that designated by terms such as 'good' were a special kind of "non-natural" property, irreducible to any facts about the world.

The third thesis alone is crucial to the project of naturalism. Are there evaluative properties that are mind-independent, objective, and irreducible to any matters of fact discoverable by empirical inquiry? It takes a robust metaphysical temperament to answer that question in the affirmative. To say No is not to reject objective ethical truth, but perhaps to allow that evaluative properties are *relative* to something about the situation and mentality of human beings. Relativity is compatible with objectivity. Moral and other evaluative judgments are plausibly regarded as relative to contingent truths about human beings; but those contingent truths, and the values that derive from them, can still be regarded as objective.

If evaluations are relative to truths about human beings, does this not mean that facts do, after all, determine values? That depends on the sorts of facts are we talking about. For those robust metaphysicians who intuit non-natural ethical properties, the facts in question are metaphysical ones, which somehow have the special capacity to make it true that I ought to do this or refrain from that. For those who do not accept the existence of such metaphysical facts, moral truths, if such exist, derive from "natural" facts about human beings and the objective world. Both

the metaphysical and the naturalist claims, however, are challenged by the first thesis I found lurking in Moore, the "is-ought" problem raised by Hume in a famous passage from the *Treatise*:

> In every system of morality I have hitherto met with, I have always remark'd that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions is and is not, I meet with no proposition that is not connected with an ought or ought not. This change is imperceptible, but is, however, of the last consequence. For as this ought or ought not expresses some new relation or affirmation, 'tis necessary that it should be observ'd and explain'd; and at the same time a reason should be given for what seems altogether inconceivable, how this new relation can be a deduction from others which are entirely different from it. (Hume 1978, III.i.3)

This passage has sometimes been described as implying that the domain of ethics, defined as that which comprises 'ought' statements, should be regarded as *autonomous* in relation to the domain of facts that are the proper study of science. As Charles Pigden (1989) has shown, the relevant sense of autonomy might be *ontological, semantic,* or *logical*. Logical autonomy is the thesis that no premises containing only statements about what *is* the case can *logically* entail a conclusion about what *ought* to be. That does not entail that there are *sui generis* moral properties. It could be that there are simply no moral properties at all: if so, then nothing can entail their presence. But recall that Hume also famously adduced reasons for believing that no premises containing statements of past states of affairs logically entail any conclusions about future states of affairs. Instead, the best we can hope for is that "custom and habit" will incline us to expect certain events as effects of other events, which we then view as causes. Why not then be content with the possibility that certain habits of mind will also, on the basis of certain experiences of fact, induce us to make certain judgements of value? I shall return to this point below.

Moore clearly intends to assert *semantic* autonomy: no evaluative statement *has the same meaning* as a statement of fact. But as Pigden rightly notes, ontological autonomy is not entailed by semantic autonomy. For while there may be no synonymy between, say, 'universal thriving' and 'good', there might be "a synthetic identity between goodness and some natural counterpart" (Pigden 1989, 128). In other words, if goodness *is* being conducive to thriving (in the constitutive sense of 'is' that is more or less equivalent to *consists in*, as in 'Water *is* H_2O,') that would not make 'goodness' synonymous with 'conducive to thriving'. It would, however, violate the purported ontological irreducibility of fact and value. Semantic autonomy, then, is not sufficient to establish the impossibility of naturalism. This is shown by Moore's own example of *yellow*, which is both natural and indefinable. Ontological irreducibility seems to be the only genuine obstacle to naturalism.

The point is worth belabouring. In a similar vein, a classic analysis of Moore's "Naturalistic Fallacy" by William Frankena argued that Moore offers us three "intuitionist" claims, which in the end boil down to one. The three claims correspond roughly to Pigden's three forms of autonomy:

(1) Ethical propositions are not deducible from non-ethical ones.
(2) Ethical characteristics are not definable in terms of non-ethical ones.
(3) Ethical characteristics are different in kind from nonethical ones. (Frankena 1939, 467).

If (3) is true, then it follows that you can't validly infer the presence of an ethical characteristic from the presence of a non-ethical one. ("They're like apples and oranges"). To do so, Frankena suggests, is to commit the "definist fallacy". That would consist in

confusing or identifying two properties,… defining one property by another, or… substituting one property for another... [in such a way] that two properties are being treated as one, and it is irrelevant, if it be the case, that one of them is natural or non-ethical and the other non-natural or ethical. (Frankena 1939, 471)

But how are we to establish that there are indeed two such disjoint sets of properties? To say X can't be defined in terms of Y just begs the question unless you've shown X and Y to be

different: "One must know that goodness is indefinable before one can argue that the definist fallacy is a fallacy." (473).

Perhaps the problem is the one that bothered Plato in the *Meno*: namely that you can only define what you already know: so any definition must be either trivial or wrong. For if I define A as B, my definition will be correct only if A and B have precisely the same meaning. But then it is trivial. If, on the other hand, I am saying that A is B when it isn't, then it's just mistaken. (But in that case, there is actually no fallacy, for a mistake is not a fallacy.)

To imply that no definition is possible would seem to be something of a reductio. Some definitions, such as 'A bachelor $=_{df}$ unmarried man' seem unimpeachable. Such definitions specify the meaning of a *word* in terms of other words that are assumed to be understood. As Frankena sums up the situation, the heart of the problem is that "[d]efinists claim to find but one characteristic where intuitionists claim to find two." (WF474). For the intuitionist, anyone defining 'good' in terms of some natural property is like a colour-blind person who sees but one hue where others see two. 'Yellow' is also indefinable: you can state no necessary and sufficient conditions for being yellow; yet yellow is a "natural" property. It admits of an ostensive definition. By contrast, you can't pick up the property 'good' from sense experience. Intuitionist objectivists such as Moore seem to think you can still define it ostensively, but only if you have a "moral sense". Moral sense, if we are to regard is as giving access to an objective world of moral fact, may not be accessible to all.

Anti-naturalists like Moore are intuitionists; but they are not the only opponents of anti-naturalism. Some, in the Kantian tradition, insist on the need for a rational foundation that owes nothing either to facts or to intuition. But if there are no *facts* that can support a moral judgement, then it must seem to follow that all moral judgements are simply unsupportable, or else that they are supported by some kind of non-facts. But what is there in the realm of non-facts that could support anything at all?

The appeal to pure reasons suggests that the relevant non-facts might be *nonempirical* facts, or what Hume called "relations of ideas". Some philosophers have thought mathematics was made up of such relations of ideas, or analytic statements, though this is rejected by Kant and by most philosophers of mathematics. Regardless of the position one takes on its analyticity, one might think, as did Plato, that our knowledge of mathematics constitutes an appropriate model for our knowledge of morality. Like Meno's slave, all of us are able to tap into a kind of knowledge that requires no support from empirical evidence. Mathematics, at least as classically conceived, proceeds on the basis of self-evident axioms. Until relatively recently, axioms were regarded as giving insight into an absolute realm of mathematical truth. After the rediscovery and reinterpretation of Giovanni Saccheri's supposed vindication of the parallels postulate, by *reductio* of its denial, we now think of mathematical theories as abstract models that require a further postulate of interpretation before we can think of them as expressing truths about the world (Giere 1984). As far as we can tell by observing objects of medium-size at medium speeds, the world is Euclidean; but at extreme distances and very large velocities, it requires us to apply a geometry of positive curvature. In the case of geometry, that decision is a pragmatic one made on the basis of empirical evidence: the original interpretation of a straight line as the path of a light ray is stipulative, but the decision to continue to use that interpretation is a pragmatic one, based on its success in generating useful predictions, even at the cost of giving up Euclidean geometry as applicable to space. Similarly, we might reinterpret the Kantian Kingdom of Ends as a purely abstract model, which may or may not be applicable to any particular community of sentient beings. Moral theories, on this view, would stand as models that can be applied to human affairs. As in the case of mathematical truths, this will sometimes result in our having to choose between two or more incompatible and apparently equally plausible models. The choice will be pragmatic there again.

If so, what might constitute the basis of such a pragmatic choice between alternative ethical systems?

Consider the claim of the Kantian categorical imperative, with its supposedly rational method of justification. I leave aside the second formulation of the imperative, grounded on the idea that rational beings should be treated as ends in themselves. That does seem to be as irreducible a moral intuition as we are likely to find, but it is not clear in what sense it can claim to be grounded in pure reason rather than intuition. The first formulation, by contrast, does claim to do just that. It requires us to conduct a thought experiment to find out if a certain "maxim" of action can be coherently universalized. Kant offers several illustrations. Only the case of lying is at all plausible, in that it appears to lead to a potential incoherence on the basis of strictly logical considerations. The universalization of lying would remove the very possibility of lying. For it would eliminate the expectation of truth, on which successful lying depends. That argument is neat; but it is also misleading. It begs the question of whether universalizability is the right thought-experiment to appeal to in the first place. Why not instead ask whether an *occasional* lie is likely to succeed? That, surely, is what the average liar will be counting on: if lies are sufficiently rare, they can count on the maintenance of the practice of truth telling. That strategy is illustrated by natural experiments, in the form of the fairly widespread phenomenon of *mimetism*. The viceroy butterfly, for example, imitates the markings of the monarch butterfly. The latter is venomous, and so his natural predators have learned (or been selected) to avoid it. By the same mechanism, those predators are induced to avoid the viceroy, despite the fact that the viceroy doesn't bother with the expense of producing venom. If monarchs became scarce or disappeared, however, the viceroy's strategy would fail. Its protection would soon vanish, together with the conditions that ensured the deterrent effect of the monarch's characteristic markings. Predators would cease to avoid butterflies sporting those markings.

In any case, universalizability is not obviously coherent on its own terms, as illustrated by the case of the policeman trained in Kantian logic: 'Move along! if everyone just stood around, nobody could get by!'. Unless carefully gerrymandered, nearly everything we do would fail the test. *What if everyone took this bus?* Consistency is a good thing, and universalizability can

usefully function as a slogan, reminding us of the rational requirement to support divergent policies with relevant differences. But though the categorical imperative is reverently taught, the uncorrupted student will dismiss this supreme principle of Reason as just dogma.

The foregoing considerations suffice to suggest that the validity of Kant's method of universalization is open to at least as much doubt as the maxims of conduct it is intended to support. That makes is subject to a broad methodological principle that any pragmatist might endorse: *there is no point in adducing dubious premises in support of a conclusion that a priori appears less doubtful than some of those premises*. In the light of that principle, the antimoralist does not need to *prove* that universalizability provides no justification for any moral rules. It is sufficient to show that universalizability as a principle is at least as doubtful as any of the specific principles of conduct it purports to justify.

2. Naturalization and its critics.

None of what I have said so far shows exonerates inferences that leap across the fact-value gap. Some recent attempts to do just that have met with powerful objections. In a recent book, Pat Churchland argues that we can best approach an understanding of morality by looking for the sources of moral attitudes in the brain. Churchland points out that Moore's own alternative to a naturalistic characterisation of the good appeals to unanalysable and irreducible intuitions. "Intuitions, after all, are products of the brain", Churchland remarks; "they are not miraculous channels to the Truth. They are generated in some way by nervous systems; they are undoubtedly dependent on experience and cultural practices, however hidden from consciousness the causes may be." (Churchland 2012, 190). Thus the Intuitionist is confronted with a self undermining paradox. She insists that we are not to look to facts about the brain (or any other facts) in our attempt to justify our moral intuitions; but the very moral intuitions that are taken to be authoritative are themselves products of the brain, which in themselves carry no guarantee of infallibility.

In opposition to Churchland's claim that "we regularly figure out what to do based on the facts of the case", however, Tom Hurka has insisted that her strategy has not bridged the is-ought gap. Her examples of fact-driven norms are irrelevant, Hurka claims, because decisions based on facts are determined by hypothetical imperatives, and their force relies on our assenting to a major premise that the scientific facts alone fail to supply (Hurka 2012). Morality, by contrast, is about Categorical Imperatives. It's not enough, Hurka writes, to point out that decent people would be against torture. For when "decent people say 'it's wrong to torture' they don't mean 'avoid torturing but only when avoiding it will further some goal you have; otherwise feel free to torture'." Thus Churchland's examples of "mundane moral thought... determined from a set of facts" (such as doing something your neighbour needs done) presupposes "a massive unstated assumption, namely [for example] that I want my neighbours' orchard not be destroyed". And again, as he puts it in a striking phrase, "because moral judgements are categorical, my judgement that lying is wrong implies that it would be wrong of me to lie even if I didn't think lying is wrong and had no negative attitude towards it."

The observation is surely correct. We use moral language, and values-language more generally, to express evaluative beliefs which we regard as objective. And the mark of objectivity is precisely this: that we think something would hold even if you didn't believe it.[1]

Hurka's attack fails. It is his observation, not the one he criticizes in Churchland, which is irrelevant: for our conviction that *if something is true, it would be true even if we didn't believe it* is not peculiar to the ethical realm. It is something we are prepared to assert of anything we believe to be objectively true. The objective fact alluded to can be a fact about value, even if values are held to be response-dependent properties. Secondary properties (such as Moore's example of 'yellow') are response-dependent in this sense: to be yellow is not to have a determinate property defined in

_____

[1] It is said that Niels Bohr had a horseshoe hanging on his door, "for luck." When someone asked, "Surely you don't believe in that superstition?" Bohr answered "Of course not; but apparently it works even if you don't believe in it."

terms of any specific light frequency, but to have the capacity to produce, in normal viewers under normal circumstances, the impression of yellow. Since circumstances can be abnormal, this allows for some things to appear yellow when they are not, and for some yellow things not to seem yellow. The property *yellow* is both objective and relative. Similarly, ethical properties can be regarded as response-dependent. They may be both objective and relative.

But what are they relative to?

One answer is inspired by Aristotle, and advocated by a number of contemporary philosophers who describe themselves as "virtue theorists." Virtue theory posits a substantive equation between the good and thriving as a human being. As such, it does not require us to believe in extra-human, absolute and objective moral truths, but it does seem committed to the existence of a universal human nature. Oddly enough, the most historically influential form of naturalism is actually that associated with supernaturalism, in the form of the theological grounding of ethics advocated by Aquinas, who borrowed the key philosophical move from Aristotle. That move, which is at the core of "Natural Law" theory, is something of a bait-and-switch proposition, playing with the ambiguities in both 'nature' and 'law'. It begins by relying for its normative force on making sense of the idea that certain things that actually occur in nature are "unnatural." The bait is the promise of looking to nature for a revelation of what nature itself "intends", as if it were seeking to uncover laws in the sense in which that term is understood in science. The switch occurs when encountering exceptions to the alleged law: instead of regarding these as falsifications, it condemns them as normatively unacceptable on the basis of their incompatibility with that "law" — thus switching, in effect, from the scientific to the legislative use of 'law'. The natural law approach promotes statistical norms to the status of moral norms, in that it relies on Aristotle' s criterion—namely that it should happen "always or for the most part."— for detecting what nature intends. If we could assume that species characteristics remain unchanged forever, that might be a defensible criterion. But in the light of the fact of evolution, it is obvious that, from the perspective of what could then be expected "always or for the most part", everyone of our ancestors who

brought us a step closer on the road from being unicellular organisms to being humans must necessarily have been a freak of nature: if all our ancestors had been normal, we would be unicellular organisms.

There have been a number of modern attempts at grounding ethics in Evolutionary Theory (Thompson 1995); but none has been at all convincing. Surely, however, our norms cannot reasonably be elaborated in wilful ignorance of the facts of nature. Morality has no need to forbid us to do what we cannot do. Neither can it require the impossible. Historically, however, it has tried to do both of these things, as attested by medieval trials of animals for murder or theft, or by laws against consorting with the devil. In order to make sure we are not still doing either, we must seek the guidance of nature. But there is a minimalist and a maximalist way of interpreting what such guidance might involve. The maximalist account endeavours to find novel and relevant facts as these are revealed to us by evolutionary theory, psychology and brain science. There are many working examples of how useful these could be if they could only persuade politicians to take them into account. Doris Lessing, in a little book entitled *Prisons We Choose To Live Inside,* pointed out that experiments such as those of Stanley Milgram on obedience to authority, although they are well known of anyone who has taken a course in psychology, might have the effect of inoculating ordinary people against the disposition to excessive obedience if they were to be disseminated among the general population (Lessing 1991). Other examples are easy to find: what is known about the nature of addiction; the actual effect of drug use, when separated from the effects of its prohibition; the effects of incarceration for young people; and many other related facts would amply suffice to demonstrate the evil of the War on Drugs. From the philosophical point of view, however, the argument premised on facts such as these remains an enthymeme, and rests on an evaluative major premise that is precisely the one attacked by anti-naturalists. That missing premise consists in the identification that a realist ethics rejects; the identification of *thriving and happiness* with *good,* and *pain and coercion without justification* with what is *inherently bad*.

In Aristote's world, the premise in question is based on a conception of what a person should be that is simply taken for granted by common consensus, transcending the vagaries of individual preference: "If, then, there is some end of the things we do, which we desire for its own sake … clearly this must be the good and the chief good": so we are told at the beginning of the *Nicomachean Ethics.* Thomistic Natural Law ethics, as already noted, is thought of as essentially theological in its foundations, but insofar as it simply adopts Aristotle's view, God doesn't function essentially in the argument. (He is more like a figurehead monarch in a constitutional monarchy.) What does the work is the Aristotelian notion of a non-intentional teleology embedded in nature itself, possibly as a whole (though this is controversial (Broadie 1990)), but certainly at the level of individual organisms, regarded as members of a species with a fixed nature. If teleology is inherent in nature itself, then it would seem that we can derive at least *some* normative statements from those teleological natural facts.

Moore was surely right in claiming that, given any proposed definition from some evaluative term, such as 'good', it is always possible to raise the question of whether the definiens is indeed good. But this is true of any definition. Take the definition of a point as the intersection of two straight lines; it is surely possible to ask whether the intersection of two straight lines is indeed a point. The answer that we give in this case would simply be, well, yes it is. And the proponent might add, if pressed: "if you have a better definition to offer, please do so." And what is noteworthy is that in this exchange, which started out being about a stipulative definition that had nothing to do with ethics, we are now faced with a term, 'better', which itself now brings the conversation into the normative sphere. In the case of mathematics, however, we find it easier to explain what it means for a definition to be preferable to some other. We can appeal to the nature of mathematics, considered as something like a game, which as such has an inherent teleology, in relation to which some definition can be deemed more or less conducive to the game's intrinsic goals. This is what Carnap would have described as a move from the semantic to the pragmatic level. When we move to the pragmatic level in Ethics, we are arguably presupposing the

meaningfulness of teleological statements, but we are not committed thereby to any particular normative claims. Those come only when we try to justify the claiming that happiness, thriving, etc. are good or that pain is bad.

3. Mill's Ploy: the desired and the desirable.

From the point of view of common-sense, facts such as the felt aversion to pain and the desired nature of pleasure appear to have a special status: they are facts, but they seem to exemplify something very close to the inherent teleology that Aristotle found in nature. They challenge the anti-naturalist to explain why they might *not* be a suitable guide to evaluative or ethical judgment. That fact underlies the supposed "proof" of the Principle of Utility for which Stuart Mill's has been criticised: Just as being seen shows something to be visible, he asserted, so 'the sole evidence it is possible to produce that anything is desirable is that people do actually desire it.' (Mill 1991). The claim exploits an ambiguity in the suffix '-able' or 'ible', which indicates worthiness in 'desirable' but signals mere possibility in 'visible'. Semantically, the criticism is correct: the two uses of this suffix are indeed distinct. Nevertheless, I would contend that Mill was right. If indeed no fact can provide evidence for value, then there cannot be any evidence for desirability *at all*. Trivially, then, there can be no *better* evidence for desirability. Mill doesn't have to claim that desirability *follow*s from desire. Rather, the actual presence of desire seems to be as good a reason as we can get. It owes that status to the fact that desire, like pleasure and the avoidance of pain, is a *why-stopper*: when we question someone about their reasons for wanting something, it makes no sense to persist when they have mentioned pleasure or the avoidance of pain — unless there are countervailing considerations. From the first person point of view, to desire something is normally to find it desirable. The desire might be endorsed, or might be repudiated by a second-order desire; but once one has assimilated the fact that there can be no other foundational fact, Mills inference seems to be reasonable even though it is sanctioned neither by logic nor semantics.

If you still yearn for a more compelling justification, I suggest you might again recall Hume's contention that inductive inference cannot be provided with any deductive justification. The practice of inductive inference is, in virtue of the way our minds are constructed, just something we do. Furthermore, as Nelson Goodman has pointed out, although deductive inference seldom seems problematic in the same way, one can do no better for the justification of deductive rules of inference. The best we can do for the justification of deduction is pragmatic and looks "flagrantly circular": "A rule is amended if it yields an inference we are unwilling to accept. An inference is rejected if it violates a rule we are unwilling to amend." (Goodman 1983, 64). This is formally similar to the quest for "reflective equilibrium" that John Rawls recommends as the test for ethical practice and principle. In the absence of a consensus on foundations, nothing is going to be either required or possible in ethical reasoning, other than the pragmatic endorsement of reflective equilibrium. Although these names are seldom brought together in the same sentence, I venture to suggest that this line of thought is supported by an unlikely alliance of Nietzsche, Hume, Goodman and Rawls. Rawls's appeal to reflective equilibrium is of a piece with Nietzsche's contention that we should forget about the justification of ethics and attend instead to its genealogy; with Hume's reduction of our inductive knowledge of cause and effect to "custom and habit"; and with Goodman's characterization of the predicates we commonly use as owing their privileged position merely to the fact that they are "entrenched" in existing projective practice.

The question raised by the practice of seeking reflective equilibrium is: what are the factors that are weighed in order to arrive at equilibrium? An increasingly compelling body of empirical research seems to show that genuine moral judgments *consist* in part in emotion.

4. The Five or Six Domains of Ethics.

Much evidence for this claim has been adduced by Jonathan Haidt and his collaborators. For most of humanity, morality involves five relatively distinct domains of intense concern, grounded in specific ranges of emotional dispositions (Haidt and Joseph 2007; Haidt and

15

Bjorklund 2008).[2] These are *harm avoidance*, *fairness*, *community-loyalty, authority-hierarchy,* and *purity*. About each of those emotional ranges we can tell a plausible evolutionary story. The resulting capacities for response generate widely diverse emotions that enforce divergent cultural norms. The first two domains—avoidance of harm, and fairness or reciprocity—are those stressed in modern philosophical ethics deriving from Mill and Rawls. Both afford clear illustrations of the innate nature of their supporting dispositions as well as the variety of ways in which they are actualized to fit into different cultural expectations.

What counts as real harm varies across cultures, which makes the exact scope of harm avoidance variable or unclear. Nevertheless, the emotional building block of the desire to avoid harm appears emerges in very young infants, and thus appears to be independent of cultural context. As obstetric ward nurses well know, a single shrieking infant can cue a chorus. Beyond infancy, empathy is modulated by numerous cognitive and cultural factors: deserved punishment or a medical procedure won't elicit as much compassion as an accidental injury (Vignemont and Singer 2006). And of course you won't feel empathy for something or someone outside the group to which your concern extends.[3]

Fairness and reciprocity illustrate the same mix of innate dispositions refined by culture into divergent practices. Aristotle unexceptionably defined justice as giving to each what they deserve. But who deserves what? Different societies obviously give different answers; but here again one can glimpse a primitive emotional 'module' which seems to motivate a simple form of concern for fairness. In the 'Ultimatum Game', one participant is given a sum of which they can arbitrarily offer any portion to a second participant. If the latter accepts, the money is allocated accordingly. If the offer is rejected, neither gets anything. In a wide variety of cultures, offers

---

[2]In more recent work, Haidt has added a sixth domain, Liberty (Haidt 2012). Others, notably Richard Shweder (2000), prefer to carve the ethical domain into three components.
[3]Describing revenge in a tribal culture, Jared Diamond (2008) describes a man who had caused an enemy's lifelong paralysis as perfectly happy—at least until a realignment of tribal alliances allows the victim to be reclassified as friendly.

below 25% to 40% have been found to trigger frequent rejection, although rejection violates economic rationality. Direct inspection of the players' brains at the moment of decision confirms that rejection has an emotional cost. It seems to be driven by a desire to punish an unfair offer. In this, humans may be unique: for chimpanzees are apparently immune to the temptation of costly spite (Jensen, Hare, Call, et al. 2006); but other primates do display something like an emotional response to unfairness, distinct from their responses to others' harm.

The other three domains, *community*, *authority*, and *purity*, are also associated with specific ranges of emotions that enforce them. In each case, plausible homologues can be discerned in other primates. It is also easy to form plausible hypotheses about the evolutionary origins of modules designed to organize responses to characteristic life concerns. In many parts of the world, for example, what Westerners disparage as nepotism is regarded as respect or loyalty to Community. Preferring a complete stranger to a relative is deemed heartless and disloyal. The underlying emotional disposition here is predicted by kin-selection theory, resting on the fact that natural selection promotes not the welfare of individuals, but the replication of heritable patterns. Hence the attachment to kin, which can sometimes motivate individual sacrifice. Loyalty morphs in many directions, from innocent pride in a home team to suicide bombing.

The emotional dispositions that support *authority-hierarchy* can plausibly be traced to dominance hierarchies observed in other social mammals—or, for that matter, in every office and every kindergarten. But authority is in tension with liberty. As Robert Paul Wolff has forcefully argued, the very notion of authority conflicts with Kantian autonomy (Wolff 1970). That notion transcends the distinctions I have drawn between different domains: whatever your standards of good and right, what matters to a Kantian is whether your act stems from an autonomous 'good will', in a sense that escapes any merely empirical tests of freedom or motive but belongs to the 'noumenal' world. It is therefore rather more difficult to find an animal model for it. In the noumenal world, everything is what it is absolutely and in itself, not relative to the perspective of someone experiencing it. By the same token, it is all radically unknowable, except insofar as the

noumenal will pierces the metaphysical membrane that hides it from view. But once you get over the awe this idea inspires, it is hard not to view it as a hypertrophic manifestation of the cult of *purity*. In Kant's Kingdom of Ends, as in the realm of Platonic forms, we are insulated from the coarseness of fleshly life. The impression is not lessened when one reads Kant's rantings about the abomination that is sex, which can be redeemed only by a contract of reciprocal genital ownership called marriage (Kant 1997). The emotional ground of purity is *disgust*, which we can assume arose by natural selection to protect us from harmful parasites and germs, but which, like the other emotions discussed, takes widely different forms. Extended to metaphorical pollution, purity takes on religious dimensions, which subsist in many supposedly secular invocations of the 'sanctity of life' and 'human dignity' (Ogien 2007).

Among the more unsettling and curious findings that have emerged from Haidt's research is the fact that emotional responses can influence moral judgments about matters to which the emotional response is completely irrelevant. In experiments conducted by Thalia Wheatley and Jonathan Haidt, for example, arbitrary words ('take' and 'ever') were first associated with disgust in hypnotized subjects. In a later phase, subjects (who had no recollection of the hypnotic episode) read a simple anecdote, in one of two versions differing only in that only one contained the words 'take' or 'ever'. The result was that the action described in the story was judged more severely by those whose versions contained the tainted words. Even about entirely innocent protagonists the subjects confabulated some reason for reproach, saying, for example, 'It just seems like he's up to something.' (Wheatley and Haidt 2005). If some value judgments can be influenced by irrelevant emotions, how can we distinguish between legitimate and illegitimate ethical judgments?

The answer, I have suggested, is to confront all our judgments and emotional responses to one another in the hope of arriving at a reflective equilibrium. But confrontations among emotions are not resolved by a simple dynamic of intensity. Emotions are moulded into widely divergent convictions, action tendencies, and normative judgments by individual reflection as well as by the less transparent forces that shape cultural change. Among those forces is the power of words:

18

conversation, debate, rhetoric, and argument, all of which are bathed in passion, but all of which still allow for rational debate.

5. Antimoralism and moralizing.

The quest for reflective equilibrium, in this inclusive sense, takes place within individuals as well as among them. Anything can be disputed. We cannot assume, in particular, that ethical values trump all others. On the contrary: any proposed way of marking out the domain of morality will itself appeal to a non-moral value—aesthetic, political, or even religious. Asserting the supremacy of aesthetics was the heroic route taken by Oscar Wilde, Baudelaire, Huysmans, and perhaps de Sade. Taking the religious to trump morality characterizes fundamentalist Christians and Muslims. Thus there may be religious, political or aesthetic grounds for restricting the scope of morality to some of the domains of traditional morality, or on the contrary for including them.

The resulting debate is inevitably political. Haidt himself has stressed that there is a major ethical divide between the conception of ethics favoured by modern liberal societies and those recognized by traditional societies, or by conservatives in Western societies. As a member of one of these modern liberal societies, it seems to me obvious that only the first two components identified by Haidt (with the addition of liberty, which he has belatedly included in his list) deserves the sort of priority in the consideration of factors relevant to a decision that can probably properly be called moral. But if I were engaged in a debate with a Muslim or Christian religious fundamentalist, or with a member of a traditional tribe, how could I possibly find arguments that might carry conviction? I myself, I would have to concede, would in effect be *moralizing* if I were to plead that the avoidance of harm and the ideal of justice should take precedence in all cases over religious superstition, arbitrary sexual norms, tribal loyalty, and respect for traditional hierarchies. In such a conversation, the subjective intensity of the contending convictions, precisely because they are grounded in emotional states, and specifically in the emotion of disgust, would seem inevitably to condemn us to an insoluble conflict of basic values. Is there any way in which some

degree of rationality might be introduced into such a conversation, without simply resorting, to make my own case, to just the sort of moralizing that that the anti-moralist deplores?

In traditional morality, moralizing seems to be involved in the imposition of values of purity, authority and loyalty which to the liberal mindset seem to belong to archaic conceptions of social life. How can my own position, expressed in just such a judgment as is contained in the previous sentence, be distinguished from moralizing?

The beginning of an answer might lie in pointing out that I don't first have to prove that my values are specifically moral ones. On the contrary: non-moral reasons can be adduced for ceasing to count norms of purity as moral. Norms of etiquette can be reassuringly stable if they are aimed at avoiding disgust (Nichols 2002). But there is no reason to lend them the heft and mystique of morality. Norms relating to community and authority can similarly be scaled down: the obligation to pay taxes and obey the law needs no support from moral sanctions. Facts such as those cited by Harris and Churchland concern the neural underpinnings of psychological dispositions and about the effects on individual thriving of different social arrangements. They also include the inescapable fact of human diversity in inclinations, dispositions and sources of satisfaction. All are ultimately rooted in the inherent teleology of desire, and are legitimately brought into the discussion on a par with prior intuitions about the different traditional domains of morality. In the resulting confrontation, it may well come to seem obvious that we should restrict the scope of morality itself, construed inclusively as comprising all six domains, in order that it may more closely promote human thriving and happiness in all its forms.

That suggestion will persuade only inasmuch as it arouses the emotion of approval. But not all emotions are of equal value to the individuals that experience them. If they stem from 'adaptations', they were shaped, like parasites or viruses, by biological and social processes blind to individual interests. The ecology of nature gets rid of most species sooner or later (except, it seems, for parasites and viruses). The social ecology of moral persons, by contrast, may allow us to survive if it is sustained and modified by talking, reasoning, and reconfiguring our perspectives.

The resulting edifices of thought and practice can be consciously designed to enhance our capacity to respond emotionally, at a meta-level of emotional approval or disapproval, to the enhancement of our first-level emotions; to select what is to count as worthy of being selected; and to promote the possibility of multiplying human possibilities.

Among the enhanced possibilities afforded by language and sociality is another set of values, the aesthetic. These are equally rooted in emotions, and they have sometimes claimed to compete on equal terms with morality. Human imagination serves the practical ends of planning, but it also affords us the capacity to attend to qualitative nuances the value of which cannot be reduced to the simple bivalence of good and bad. In the aesthetic attitude, we can focus not on the practical guidance for which our emotional repertoires evolved, but on their intrinsic quality.

The aesthetic stance comes closest to what Aristotle characterized as 'divine', namely pure contemplation. Insofar as it is potentially in conflict with morality, such an aesthetic stance is closer to antimoralism than to morality itself. This can be underlined by drawing two contrasts between art and religion. First, art affords imaginative illusions, entertained as such; most religions, by contrast, insist on delusive commitments of belief, enforced with meretricious expectations of rewards and punishments. Second, aesthetic contemplation endorses the possibility of conflicting yet equally authentic values. Monotheistic religions must reject this, being committed to the view that all true values are compatible since they are united in God. The very structure of our emotional life, made manifest in classical tragedy, attests to the fact that conflicting values can be equally real. By expanding the scope of relevant reasons to criticize the hegemony of morality, that consideration may mitigate, at least to some extent, the moralizing tone of antimoralism.

References

Blackwood, Russell. 2010. "Book Review: Sam Harris's The Moral Landscape." *Journal of Evolution and Technology* 21(2): 53–62.

Broadie, Sarah. 1990. "Nature and Craft in Aristotelian Teleology." In *Biologie, Logique et Métaphysique Chez Aristote*, ed. Daniel Devereux and Pierre Pellegrin. Paris: CNRS.

Churchland, Patricia Smith. 2012. *Braintrust: What Neuroscience Tells Us About Morality*. Princeton: Princeton University Press.

Diamond, Jared. 2008. "Vengeance is Ours: What Can Tribal Societies Tell Us About Our Need to Get Even*?" New Yorker*, 1921 April. Accessed online on 2014/02/11 at http://www.newyorker.com/reporting/2008/04/21/ 080421fa_fact_diamond?currentPage=all.

Frankena, Wlliam K. 1939. "The Naturalistic Fallacy." *Mind* 48(192): 464–77.

Giere, Ronald. [1979] 1984. *Understanding Scientific Reasoning*. 2_ ed. New York: CBS College Publishing.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon.

Haidt, Jonathan, and Fredrik Bjorklund. 2008. "Social Intuitionists Answer Six Questions About Moral Psychology." In *Moral Psychology, Vol. 2*, Walter Sinnot-Armstrong. Cambridge, MA: MIT.

Haidt, Jonathan, and Craig Joseph. 2007. "The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules." In *The Innate Mind: Foundations and the Future, Vol. 3*, Peter Carruthers, Stephen. Laurence, and Stephen Stich. Oxford; New York: Oxford University Press.

Harris, Sam. 2011. *The Moral Landscape: How Science Can Determine Human Values*. New York: Free Press.

Hume, David. [1888] 1978. *A Treatise of Human Nature*. 2d ed. Ed. L. A. Selby-Bigge. Revised by & notes by P. H. Nidditch. Oxford: Oxford University Press, Clarendon.

Hurka, Tom. 2012. "Churchland's Normative Arguments." Author Meets Critics Session on Patricia S. Churchland, Braintrust: What Neuroscience Tells us About Morality. American Philosophical Association Central Division Meeting. Chicago, February.

Jensen, K., B Hare, J. Call, and M. Tomasello. 2006. "What's in It for Me? Self-Regard Precludes Altruism and Spite in Chimpanzees." *Proceedings Biological Science B: Biological Sciences* 273(1589, 22 April): 1013–21.

Kant, Immanuel. 1997. "Duties Towards the Body in Respect of the Sexual Impulse." In *Lectures on Ethics*, trans. Louis Infield. Cambridge: Cambridge University Press.

Lessing, Doris. 1991. *Prisons We Choose to Live Inside*. Toronto: Anansi.

Mill, John Stuart. 1991. Utilitarianism. Vol. 10 of Collected Works of John Stuart Mill, ed. John M. Robson. University of Toronto Press.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press. Available online at http://fair-use.org/g-e-moore/principia-ethica.

Ogien, Ruwen. 2007. *L'Éthique Aujourd'hui: Maximalistes et Minimalistes*. Folio essais. Paris: Gallimard.

Pigden, Charles. 1989. "Logic and the Autonomy of Ethics." *Australasian Journal of Philosophy* 67(2): 127–51.

Shweder, Richard A., Nancy C. Much, Manamohan Mahapatra, and Lawrence Park. 2000. "The 'Big Three' of Morality (Autonomy, Community, Divinity) and the 'Big Three' Explanations of Suffering." In *Why Do Men Barbacue? Recipes for Cultural Psychology*, Richard A. Shweder. New York: Routledge.

Thompson, Paul, ed. 1995. *Issues in Evolutionary Ethics*. Albany, New York: State
University of New York Press.

Wheatley, Thalia, and Jonathan Haidt. 2005. "Hypnotic Disgust Makes Moral
Judgments More Severe." *Psychological Science* 16(10): 780–84.

Wolff, Robert Paul. 1970. *In Defense of Anarchism*. New York: Harper and Row.