This review-essay was written for the Literary Review of Canada, in 1993. The LRC found the piece to technical, and it has remained unpublished.

Peter Danielson, *Artificial Morality: Virtuous Robots for Virtual Games* (London, UK: Routledge, 1992).

Anyone who attends to the discussions taking place among the well-informed surely would realize that the traditional claim of ethics, that we judge an act, motive or character to be good or bad because some actual feature or features it possesses, has fallen into disrepute. Moral deliberation was believed to result in what the Canadian philosopher Charles Taylor has called "strong evaluations," that is, in his words, "discriminations of right or wrong, better or worse, higher or lower, which are not rendered valid simply by our desires, inclinations, or choices, but rather stand independent of them and offer standards by which they can be judged." Undergirding the idea of strong evaluation is a central belief of traditional ethics (of whatever stripe), that some acts and some characters are worthy of praise or condemnation because they have or lack identifiable qualities.

Many thorny problems arise from the claim. To think thoroughly about these moral issues is to encounter questions about the ontological status of the properties that support these claims: whether they are natural or non-natural; whether they inhere in the act or character evaluated by itself or whether they are relational properties, depending, for example, on the relation between the act, motive or character that we evaluate and the cosmic order; and—perhaps most important of all—whether we learn of their presence by the senses, by reason, by intuition, or solely by revelation.

These are such important questions that the varieties of answers that one can give to them established the frameworks of the traditional schools of ethics. However, almost all agreed that actual features of acts, motives and characters provided the basis for our evaluation that they merit praise or condemnation.

There is no longer such agreement. Few now maintain that values are objective (or have an objective basis). The common currency in ethical theory now is that values are subjective and reveal facts about the individual who frames an evaluation, not about the object evaluated.

The reasons this consensus collapsed are the key to understanding the character of that form of life we call modernity. Some argue the collapse happened when sociology showed that our beliefs and values vary from culture to culture. If values are culturally relative, features of the acts, motives or characters evaluated cannot determine them, as these acts, motives or characters presumable are transcultural phenomena. Evaluation therefore must depend instead on the frame of reference the evaluate employs in making the evaluation.

I do not believe that this analysis has enucleated the core of the problem. That evaluations lost their objective ground certainly is true, and that fact, undoubtedly, is a key to the configuration of ideas, beliefs, attitudes and judgments that constitute modernity. But I do not believe that it was sociology's teaching that values are culturally relative that discredited and dismissed the traditional belief that actual features of acts, motives and characters ground our moral evaluations. Rather, it was the centrality that physics, and more particularly to the physics that developed from the late sixteenth century onward, has had in the modern paradigm. More exactly, it was the modern paradigm's giving over its conception of nature to that which undergirds physics as it has developed since the sixteenth century and finds its fullest expression in the philosophy of Thomas Hobbes. According to this physics, matter is all but featureless, having only mass and

extension. Only the movement of this nearly featureless matter has a rightful place in a scientific account. No scientist should petition to the ends that objects serve to frame an explanation for their conditions.

Such Classic virtues as loyalty and gratitude lost their standing when moderns rejected the classical belief in the objectivity of value; as this happened, Stoic virtues to returned to take the place of the more orthodox classical virtues that the early moderns had just abandoned. Self-control and equanimity became the paragon virtues exactly because they are subjective conditions that need not respond to objective conditions. To give way to anger, or even to judge events, people or objects became the cardinal sin. Within the moral framework to which the Stoic virtues belong, all condemnation is equally opprobrious, since it gives evidence of an impassioned and uncontrolled state of mind, One needs to understand how we have come to this. There is likely no better place to start than Prof. Danielson's book.

Danielson is Associate Professor of Philosophy and Senior Research Fellow in the Centre for Applied Ethics at the University of British Columbia. He offers his book as work that undertakes the task that philosophers call fundamental justification. That is, it undertakes to provide a justification for the realm of morality that does not invoke any notions specific to ethics. Most often, the justification consists in showing that the principles of reason lead one to adopt moral standards, that it is rational to act morally. Danielson, too, adopts this tack. In this he follows David Gauthier, a University of Toronto philosophy professor whose work in fundamental justification has attracted commentary from around the world and is the subject of the volume *Contractarianism and Rational Choice: Essays on Gauthier*. In his influential book, *Morals by Agreement*, Gauthier propounds a version of instrumental contractarianism, according to which moral agents agree among themselves to cooperate because they realize that they fare better by cooperating. Among Gauthier's important contributions to the theory of the instrumental contractarianism is the moral principle he calls 'constrained maximization,' according to which constrained (moral) actions must be conditional upon the cooperation of other agents to be rational. Gauthier compares the gain (utility) that accrues to a conditional maximizer with that which accrues to an amoral exploiter ('a straightforward maximizer'), an agent who always acts to ensure maximum immediate gain. His arguments show that the constrained maximizer in the end does as well or better, on his conception of utility, than amoral, straightforward maximizers.

The idea of maximizing utility is grounds for the use Gauthier and Danielson make of a model of moral behaviour. This model is the so-called prisoner's dilemma. It was given the form and significance it currently holds at that notorious centre of American militarism, the RAND Corporation. A scientific statement and analysis of prisoner's dilemma appeared in June 1952, in a RAND research memorandum by Merrill Flood that described an experiment that he and a colleague at RAND, Melvin Dresher had conducted in January 1950. Another RAND consultant, Albert Tucker, described the experiment succinctly in a letter to Melvin Dresher, describing it as "a dressed up version of the game you showed me." Here is Tucker's description of the game:

> Two men, charged with a joint violation of law, are held separately by the police.
> Each is told that
> (1) if one confesses and the other does not the former will be given a reward . . .
> and the latter will be fined . . .
> (2) if both confess, each will be fined . . .
> At the same time, each has good reason to believe that

(3) if neither confesses, both will go free.

Computer scientist and artificial intelligence researcher Douglas Hofstadter suggests another model, equivalent to the prisoner's dilemma in game-theoretical structure, is clearer.

Here is Hofstadter's model (with some changes that I have introduced to make it plausible for these post-Mulroney years): Suppose you wanted to buy a large volume of crack cocaine from your neighbour supplier, but wanted to minimize the risk of being caught. You might take a walk to nearest convenience store and wait. Lo and behold, in just a few moments he appears! Much to your delight, the businessperson offers you an especially good deal. He wants to escape the heat, he tells you, so he'll sell you his entire remaining supply at a discounted amount which he will use to flee to the city. You agree to the purchase. After all, it will take but a couple of seminars, which you are easily able to drowse through, to pay for a mighty fistful of the stuff.

But you're an anxious sort. To lessen your fears, you strike upon the idea that the two of you will agree upon a drop-off and pickup spot. Your local entrepreneur will leave a back with the cocaine in it, hidden in a garbage pail behind the local subway station. You will come along, leave your money in another garbage-pail, around at the side of the building, continue to the back, pick up the bag full of cocaine, and leave. The enterprising young businessperson will then return, pick up the money and leave.

That's fine, you think. Then you think better of your plan. You realize that the street corner businessperson might not keep his part of the promise. He might simply leave a sack of sugar instead of the cocaine, and then you would be out your salary. Well, you think, two can play this game. What is more, I can think of a no-lose strategy. What I must not do is give him my money—that's the sucker's move, and I will not play the sucker. No matter what, don't give him the money. Then, if he leaves the cocaine, great!—I get it for free. If he does not, well that's alright, since then I will have gotten nothing for nothing.

But the local merchant is a pretty sharp fellow too. He follows much the same line of reasoning as you did and concluded that, no matter what you do, he should not leave you your fortifier. Whether you leave a bag of money or not, he will be better off by leaving a bag of white sugar than by leaving a bag of cocaine, he reasons.

Thus it seems logic prevents cooperation. This is a sad state of affairs, because if the two of you were to cooperate, you would both get something that you want. Game theorists would say that the prisoner's dilemma produce outcomes that are not Pareto optimal (after the Italian economist Vilfredo Pareto)—that is, it does not produce the outcome that has maximal utility for the two parties (inasmuch as there exists another outcome that has greater utility for one party and no less utility for the other.)

This is gloomy state of affairs: moral behaviour was brought before the court of reason and was found wanting. Situations with the game-theoretical structure of the prisoner's dilemma appear to undermine the rational basis—and, given our culture's hypervaluation of reason, the *only* conceivable basis—for morality.

Thomas Hobbes based his theory of political authority on considerations not unlike those I have just set out. Game theory assumes that the players are self-interested, rational agents; Thomas Hobbes had posited that individuals in the state of nature have the same characteristics. Hobbes used a grand-scale analysis of human interaction that was not unlike that of giant prisoner's dilemma to argue that rational, self-interested agents will see that it is necessary to associate penalties with cheating at the game and not keeping one's promises—penalties that are harsh enough to make cheating irrational. Technically, in game theoretic terms, he proposed to change the utility functions of the outcomes (i.e., their place in scale of relative preferences.) What it would take to alter the player's utility function and make cheating undesirable, Hobbes concluded, is a sovereign with the power to detect and punish cheating.

Thus Hobbes justified state authority. Moral reasoning alone was not sufficient to lead us out of the nasty state of nature; we require a sovereign with the authority to punish stringently.

In the example above, we set out a situation in which you and the drug dealer engaged in a single exchange, with the knowledge that you will conduct no further transactions with him. We could develop the puzzle further, in a way that might reveal more about the structure of moral behaviour. Suppose that, the police-board, in response to the pressures of a citizens' action group, and acting at the urging of an enlightened and tolerant provincial premier, forbade harassment of our local merchants. This was quietly communicated to the merchants in neighbourhood, and so our businessperson could continue conducting his business down at the local convenience with knowledge of his impunity. You are pleased, because his stimulant had made it possible for to go on working in classrooms where students know less than nothing. You would like the street-corner entrepreneur to become your regular supplier. Despite the police-board's enlightened policy, you continue to be anxious, and want to continue with the method for making your exchange that you worked out for the initial situation. First time out, both of you figure that you must fulfill your promises (yours to leave a bag of money, his to leave a bag of real, high-potency crack cocaine) to create the basis for this mutually beneficial relation being continued. You both reason that the other will think the same way; and, believing this about him, you conclude you are not in danger of being taken for sucker if you do leave the money.

Your need for crack cocaine has increased as the quality of your students has plummeted and by now you need more every week. Each week therefore you have to remake the decision to leave the money or bag of cutup newspapers, and each week your supplier has to remake the decision whether to leave genuine cocaine or whether to leave white sugar. In the language that thinkers have developed for the analysis of the prisoner's dilemma, you must decide whether to *cooperate* (leave a bag of money) or *defect* (leave a bag of cutup newspapers).

Game theorists call the situation we have just described, in which the decision to cooperate or defect must be taken repeatedly, the iterated prisoner's dilemma (while the first situation, in which you and your local merchant engage in single transaction they refer to as noniterated prisoner's dilemma.) What strategy would a rational player adopt for the iterated prisoner's dilemma? If, for example, one month you found a bag of sugar instead of cocaine, would you simply conclude that you cannot any longer trust the good merchant who had furnished you so well until then and decide that you from now on you must leave bags of cutup newsprint? This strategy (known as massive retaliation) would surely result, sometime in the future, in the seller's decision to start to leave only bags of sugar and so you would adopt it at the cost of losing your supply. Should you then continue to leave bags full of money with the hope that you will gain the trust of the businessperson, and that he will respond by leaving bags of cocaine? If the strategy works, you attain your goal of ensuring a regular supply of what you need to face your students, but if it does not, you would be playing the sucker. Perhaps you should leave money for a set number of times in the future; if the businessperson has not restarted your supply by that number of times, you will stop leaving money.

How can one decide what strategy to adopt? Game theorists calculate the benefit that accrues to the two agents by creating a utility function—a formula that assigns points that represent the benefit or costs to the players of the strategies they adopt. The cost/benefit results are often set out in a *payoff matrix*. In the matrix for our game, when both players keep their word each gets 2 points (both get what they want, but at a cost). If neither keeps his word, neither gets what he wants and both gain 0 points. If one keeps his word, and the other does not, then one player reaps 4 points, the other, who suffers a loss with no compensating gains, loses a point.

The University of Toronto philosopher and psychologist Anatol Rapoport proposed the following strategy for the iterated prisoner's dilemma in response to a request from Robert Axelrod of the Political Science Department and the Institute for Public Policy Studies of the University of Michigan in Ann Arbor for a tournament, to be played by computer, among players adopting strategies various contributors had suggested. (Axelrod's findings appeared his book *The Evolution of Cooperation* published by Basic Books in 1984.) The strategy that Rapoport proposed was the simplest of all the programmes submitted. The algorithm was:

> Keep your promise on move 1.
>
> Thereafter, do whatever your opponent did on the previous move. (Keep your promise on move n+1 if he kept his on move n, cheat on move n+1 if he cheated on move n.)

Rapoport's player, whom game theorists usually call Tit For Tat, won the tournament. An important feature of Tit For Tat is that he is, in the technical language of game-theory *nice*, that is Tit For Tat will not cheat until the other player has cheated. *The Evolution of Cooperation* has high praise for nice players. Axelrod's summary of his results includes this sanguine comment:

> Even expert strategists from political science, sociology, economics, psychology, and mathematics made the systematic errors of being too competitive for their own good, for not forgiving enough, and too pessimistic about the responsiveness of the other side.

In short, good players would be nice (would not cheat first) and forgiving (would not respond to cheating with cheating for an unnecessarily long period.)

A second, much larger tournament, in which Tit For Tat won again, augmented the "moral" (but is it really moral?) prescription 'be nice and forgiving' with a third 'good personality trait' for our agent, viz. being provocable. An agent is provocable if s/he 'becomes angry' and responds to cheating with cheating—but not by massive, but by restrained retaliation. The trait of forgivingness [*sic*] ensures that, after an interval of retaliation, cooperation will be reestablished.

So, it had seemed the sun had come out to shine again on ethical theory. Game theorists had shown that traits such as being 'nice' and 'forgiving' were rational, and no one could really doubt that briefly retaliating for being cheated was morally justifiable. As Axelrod put in *The Evolution of Cooperation*, "Mutual co-operation can emerge in a world of egoists without central control [Hobbes' sovereign R.B.E], by starting with a cluster of individuals who rely on reciprocity."

This happy moment was not long-lived. David Gauthier pointed out, and Danielson agrees, that the strategy of Tit For Tat in the iterated prisoner's dilemma is morally neutral, for it does not require any 'moral' change in the player (in the sense of constraining his self-interest.) Like Thomas Hobbes' theory of political authority and Adam Smith's economic theories, Axelrod and Rapoport's Tit For Tat is an institutional solution that does not demand any internal, 'moral' change in the agent—does not demand any change in how the agent constrains his or her self-interest. Tit For Tat lacks moral relevance because his actions are directed simply to maximizing utility; and Tit For Tat's self-interest is not curbed or constrained in any way.

One of Gauthier's principal contributions to ethical theory was suggesting a means by which rational, self-interested agents can come to the conclusion that they ought to cooperate with their fellows even when there is no sovereign that might detect and punish them for their cheating ways, but simply by internal, 'moral' change that brings them to constrain their self-interest. Gauthier employs a different model than the iterated prisoner's dilemma, a model on which both players realize that they will likely find themselves in prisoner's dilemmas in the future when other players can respond to their previous moves or their intentions—that they likely can respond to your type, whether cheater or cooperator. He refers to this model of play between rational, self-interested individuals who can identify, at least sometimes the character of the opponents, as cooperators or defectors as an extended prisoner's dilemma.

Suppose you are a cheater. If you play with a cooperator, either she will recognize you as a cheater or she will not. If she does not recognize you, you can play with her and be sure of winning. However, if the cooperator recognizes you as a cheater, she will exclude you from playing, and you will win nothing. On the other hand, if you become a cooperator, you can play with cooperators whether they recognize you as a cooperator or not. You will then gain something, though your net gain will not be as great as it would have been had you played against the cooperator as a defector.

Gauthier shows the rationality of becoming a cooperator in the following way. Suppose, he says, the utility of not cooperating is u, of cooperating with other cooperators is u', and of cooperating with defectors is 0. Suppose, too, that $0 < u < u' < 1$ (that cheating has the highest utility, cooperating the next highest, not cooperating the next highest, and being an exploited cooperator none at all.) Suppose, further, that the probability of a defector meeting and recognizing a cooperator is p, of being able to exploit the cooperator is q (and hence the probability of meeting and being able to exploit a cooperator is pq) while the probability of a cooperator meeting and recognizing a fellow cooperator is r. The expected utility formulae for your two choices are:

(1) $u + pq(1 - u)$ for a defector and

(2) $u + pr(u - u) - q(1-p)u$ for a cooperator

From these formulae we can conclude that you have a chance for gain if you become a defector and a chance for a smaller gain and for a loss if you become a cooperator. But you cannot learn from them whether you will realize more utility if you become a defector or a cooperator: that will depend upon the particular values of u, u', q, p and r. This is as far a Gauthier's general case takes us.

Consider however the particular case in which p = 1, i.e., in which you are certain to meet only cooperators. Then formula (1) becomes:

$$u + q(1 - u)$$

and (2) becomes

$$u + r(u' - u)$$

Since u is a common, you can decide between becoming a defector (formula 1) and becoming a cooperator (formula 2) by comparing $q(1 - u)$ and $r(u' - u)$. The first (associated with becoming a defector) is smaller than the second (associated with becoming a cooperator) just in the case that

$$q / r < (u' - u) / (1 - u)$$

that is, just in the case that the ratio of the probability of defecting to the probability of cooperating is smaller than the ratio of the gain from cooperating to that from defecting.

Thus, if the probability of cooperating is high, and the chance for defecting is low, and the gain from cooperating is close to the gain from defecting, you should cooperate. On the other hand, if the probability of defecting is high compared with the probability for cooperating and the gain from defecting is high in comparison with the gain from cooperating, you should defect.

The conclusion is not sweeping. Nonetheless, Gauthier did show that when the benefits from cooperating are roughly the same as those from defecting, and when the probability of successfully cooperating exceeds that of defecting, it is in one's self-interest to cooperate. In a society in which the payoff from cooperating is not much less than the payoff from defecting and the possibility of successfully cooperating exceeds that of successfully defecting, it is rational to cooperate. Gauthier proves, finally, that it is rational (i.e., maximizes one's expected utility) to choose to live by a policy that requires one to cooperate faithfully when presented with cooperative ventures (a) that furnishes more utility than continuing on one's own does and (b) in which one has reasonable grounds for believing that other agents will remain faithful to their commitments.

Gauthier's findings lay the basis for examining cooperative games in which communication between the players is possible with a view to providing a fundamental justification for morality. This is the chore that Danielson undertakes in this volume.

This is the background needed to understanding Danielson's work (and which Danielson really should have provided.)   Clearly, if we accept the legitimacy of the programme of fundamental justification, many important consequences for morality and social theory follow from Danielson's work. The book is important enough that anyone concerned with our self-understanding as moral agents should read it; it epitomizes the most deleterious strand in contemporary social and moral theory.

Danielson adopts a method similar to that which Axelrod used for *The Evolution of Cooperation.* He constructs tournaments in which players who follow one strategy play serially against all players following different strategies and finds which players are most successful over the long haul.

Among the most charming features of the book is a series of programmes in Prolog, an impressive computer language (and my favourite) that grew out of Alain Colmauerer's work in the analysis of natural languages at L'universite d' Aix-Marseilles. The language has great appeal to those with traditional philosophical training as Colmauerer modelled it on Whitehead and Russell's predicate calculus and it comes closer than anything else I know to being an automatic theorem prover (write down your premises in the syntax of language, then ask it if various propositions follow from the premises.)

The book will provide a real chill to all those who are concerned with issues of our self-understanding. But it provides a basis for reflecting on the dysfunctions of modernity. In what remains, I wish to provide a few, all-too-sketchy comments on a small portion of the issues the book raises.

The first is the use of the computer to model exchanges between agents and to calculate the utility they provide. The end of this use of computers, as I noted above, is to provide a fundamental justification for the field of ethics. The task of providing a fundamental justification for some field of endeavour is to reduce the concepts of that field to concepts from another, more solidly established field.

But no efforts at justification can be presuppositionless. All must draw on the concepts embedded in another domain of intellectual or spiritual inquiry. The shape that the project of fundamental justification in our culture derives from our choice concerning which field we will use to justify morality (or from what field we will draw the the terms into which we will reduce the terms of morality or what ever field we strive to justify.) Moderns draw these concepts from logic or, more generally, the fundamental disciplines of rational inquiry.

Hence Danielson wants to show that it is rational to be moral; the form the proof takes is showing that moral agents can “successfully solve social problems that amoral agents cannot solve. The obvious way to test [this] this claim is to build worlds with social problems and see if moral agents are differentially successful in them.” (p. 4).

This method has several striking features. The first is the supposition that morality stands in need of fundamental justification while the principles of rational inquiry do not—as though the principles of rational inquiry were in some sense more primitive than the principles of reason. This already presumes ideas about morality that are far from settled. It could be, for example, that morality relies on the direct intuition of a simple property, the goodness of an act, that cannot be resolved into anything simpler. The corresponding normative proposition in the domain of aesthetics is often offered as an item of common wisdom, and there is no evident reason we should reject the ethical version out of hand. There have been ethical thinkers who have argued for it, as Danielson realizes.

The second is that the enterprise has a dual basis. One is Danielson’s conception of a moral agent. His is a startlingly empty conception of moral agency; this conception is a key to Danielson’s belief that his experiments with software robots might have something interesting to tell us about human morality or about interrelation among humans. Danielson describes his agents as “sets (generators) of actions.” (p. 64) The agents he examines have no character except the capacity to follow rules and the desire to maximize benefit to themselves. He does not even claim that human nature is inherently selfish. That his robots are simply self-interested benefit maximizers is not meant to suggest anything about the constitution of real human agents. It is a methodological hypothesis. The assumption that agents are simply self-interested provides the most difficult case for establishing the viability of cooperation. If he can prove that for this case, presumably he has proved it for all possible cases. His position, then, is not that agents are selfish. Though he models moral behaviour with “benefit maximizers,” he conceives of the model as hypothetical and uses to found a method that one might call hypothetical egoism. As he puts it, “I ask, if agents were selfish, why and how could they become morally constrained? “

The idea that morality is essentially constraint—control of the passions by reason—was common in the classical period. But that era had a considerable different conception of reason than does our own. For the classical philosophers, as for Spinoza, the order of reason (thinking) corresponds to the order of being. The person who had seen the Good would be able to derive all truths from its Being, following a line of deduction that corresponds exactly to the manner in which existents of every order of being derive from the Good. The assertions that classical philosophy offers about the proper role of reason in human life follow from descriptive propositions about the order of reality. Danielson’s do not. This difference, we shall see, is damaging to Danielson’s theory.

Danielson’s description of agents as “sets (generators) of acts” has further difficulty. Admittedly he offers this as descriptor of his software robots. Even so, this comment puts him on the horns of a dilemma. Either he believes that important disanalogies between his software robots and moral agents, or he does not. If he does, he must explain why his experiments with software robots have any relevance to discussions of what humans owe one another in their interactions, and he does not do this. If, however, he does not think there are important disanalogies, then his conception of the relations between morals rules and the nature of human being is wanting.

The enterprise of justifying morality has its beginnings in the Enlightenment (especially the Scottish Enlightenment.) For classical philosophy, the proper starting-point for ethical inquiry was the relation between human nature and virtue. George Grant pointed out insistently that classical philosophy maintains that human nature fits human being for specific ends and that the goodness of human being consists in executing well the tasks for which human nature fits human beings. It is just the same for human beings as it is for every existent. Each existent has a purpose that relates to its nature. The purpose of a knife is to cut. The purpose of a physician is to promote health. An existent's virtue depends upon its purpose. A good knife is a knife that does well what a knife is to do, that is to cut. A good physician is a physician who does well what a physician is to do, that is to promote health. So too human nature makes human being fitted to a particular sort of life, and the particular virtue of human beings depends on the life that human nature makes them fitted to live. The feature that distinguishes human being from other beings is the capacity to reason, and so the life to which humans are suited to living is a life of reason. Human goodness consists in living a life devoted to rationality, in which reason will control one's actions; for Aristotle this meant avoiding extremes and seeking a middle course between them.

That the distinctive trait of human nature is its capacity for reason was crucial in Aristotle's argument. It established a basis for morality in matters of fact. The Enlightenment rejected these claims, and proposed that moral propositions cannot be derived from statements that describe the population of the world. When moral statements lost their bases in matters of fact, the nature of moral thinking changed, for moral arguments become interminable.

In Danielson's argument we see the continuation of the classical philosophy's claim that reason must govern action. Thus, he proposes that the identifying trait of morality is "impartial self-constraint." Classical Greek philosophers such as Plato and Aristotle also thought that reasons's mastery of the self was the identifying trait of morality. However, their belief rested on a proposition about the nature of human nature—on what they considered a fact about human being. For the classical philosophers, the order of reality generally, and more specifically, the nature of human being grounded the proposition, that reason should control the passions. With the repudiation of the belief that values have a basis in the order of reality, that proposition has lost its ground, and left such assumptions as Danielson's about the essential nature of morality baseless. Alasdair MacIntyre's work, *After Virtue*, shows that the repudiation of the belief that the values have an objective ground in the order of reality makes moral controversy unresolvable.

Danielson recognizes the difficulty his position lands him in, for he makes this badgering remark about his programme of fundamental justification:

> Mine is a counter-factual argument that attempts to establish a connection between two seemingly independent and opposed realms, the rational and the moral. A less misleading term [than fundamental justification] would be 'reductive justification.' I seek to reduce (some instances of) the question: why be moral? to questions about rationality. This makes my argument fundamental. I do not go further and seek a foundation by attempting to answer the question: why be rational? The point of these premises is to block regress of this sort. All arguments are relative to assumptions: one must start somewhere. Artificial Morality begins with the assumption that amoral rationality matters.

There is considerable bravado behind the statement, but its coercive tone masks a key question namely, whether the life of reason is a good life for human beings, or whether human beings are fitted for another, perhaps a life of contemplation and prayer. Danielson rules questions about the place that reason should have in human life out of court by declaring that arguments must have a stop, and this is good place to halt.

Danielson denies that human beings are fitted for any sort of life (at least in any way that has ethical implications.) He frames his opposition in terms that are, by now, the orthodoxy, by petitioning to Darwinianism (the same basis that advocates of monetarism and laissez-fair market-driven policies invoke.) He quotes Steven Jay Gould's description of Darwin's practices and likens those practices with his own.

> Darwin advocated a natural and testable theory [cf. Danielson's experiments with robots R.B.E] based on immediate interactions among individuals (his opponents considered it heartlessly mechanistic [cf. people's attitudes towards use of robots to model human behaviour. R.B.E]. . . the balance and order of nature does not arise from a higher, external (divine) control, or from the existence of laws operating directly upon the whole, but from struggle among individuals for their own benefit.

To this Danielson appends the comment "I find Darwinianism profoundly liberating, both intellectually and politically, as it frees us from oversimplified models of central (divine or political) control of our affairs.

Really, Danielson's work, as he himself describes it, is an inquiry into "how it can be useful *for me*, the agent to constrain myself." (p. 42, emphasis his!) But, we must ask, are these prudential considerations really ethical considerations. When we say that an act is good, do we really mean that it will promote the greatest benefit for the agent in the end? (p. 42) More likely, virtues are not instrumental, as Gauthier and Danielson have it—they do not conduce to the good life but constitute it.

Danielson's conception of utility (or benefit or payoff) is just as empty as his conception of moral agents. He assigns quantitative value to certain results, but he does not—and I think cannot—acknowledge that benefit is not a single, unitary notion. There are many things that benefit might be, and it is difficult to see how, in his terms, we could compare them and assign them values. If he does not provide us with a means of assigning values to outcomes of actions, he does not show us how we might compare the benefits associated with the various outcomes. The different benefits an action produces (among which might be the pleasure of cooperating or the pride in doing something on one's own, despite other individuals' rules) might be utterly incomparable.

Treating utility as a single, unitary notion turns an ethical programme into an homogenizing one. Doing so suggests that the good that human beings seek is everywhere and always alike and that all agents in the same situation would make identical calculations of utility. The image of the moral considerations that Danielson offers is that of a vast, impersonal machine designed to ensure the maximization of utility.

William Godwin realized the universalistic implications of utilitarian consequentialism. He averred that, discovering a burning house, one certainly must save a gifted prospective benefactor of humanity (such as the good Archbishop Fénelon) and leave the less gifted behind, even if they were one's mother, one's wife or one's child.

> Pure unadulterated justice . . . would still have preferred that which was most valuable . . . What magic is there in the pronoun "my" to overturn the decision of everlasting truth? My wife or my mother may be a fool or a prostitute, malicious, lying or dishonest. If they be, of what consequence is it that they are mine?

And the orthodoxy wants us to believe that Grant's statement that love begins with love of one's own is a morally offensive claim framed to justify the ugly doctrine of nationalism! I'll say, and I'll believe it as much as that other dogma associated with positions like Danielson, that impartiality is an essential trait of moral decisions!   Perhaps professors really are that high-minded, but those of us who live on earth are not capable of such lofty and inhumane impartiality.

What Godwin's example points out is the importance of considering the social contexts of actions and the links that agents have to others around them. I do not believe that Danielson's theory can be adapted to accommodate such considerations, not even by introducing the establishment and maintenance of personal bonds as part of the utility that our actions produce. Danielson portrays ethical decision-making as an action of a solitary agent who strives to maximize his or her utility. The real problem with this is that moral agents are not Solitary Choosers, but beings that only come to be through their relation with others. Considerations of maximizing one's own interest alone are too narrow to have explanatory value in ethical theory.

The understanding of impartiality that so frequently goes along with ethical theories of Danielson's ilk too had its beginnings in the Enlightenment. The exemplar of the Scottish Enlightenment, David Hume, justified his reduction of all motives to considerations of utility by claiming that such simplification conforms to the scientific method and "was Newton's chief rule in philosophizing." Philosophers such as Newton and Hobbes sought that single notion that might introduce scientific clarity into ethical discussion.

So philosophers of the Enlightenment tried to find a means for showing that moral decision-making is really a matter of calculation—or, we might say, computation. This conception of reason has cut us off from all knowledge of the Good, without which morality could not be anything but calculating. As importantly, it obfuscated the complexity of the social context in which we make moral decisions, and the relevance of that context to the rules that implicitly guide our actions.

George Grant's darkest teaching concerning technology is that technology is not an array of neutral implements set out before us, that might use well or badly; technology has refashioned human being, as its being has penetrating into our most inward recesses. Nothing could better reveal the truth of Grant's dark teaching than this sad, inhumane book that can understand reason only as instrumental reason.

But it is a good primer on modernity.