# 2006-507: THE ETHICS ENGINE: A MATHEMATICAL APPROACH FOR MOTIVATING ENGINEERING ETHICS DISCUSSION

**Gene Moriarty, San Jose State University**

# The Ethics Engine: A Mathematical Approach for Motivating Engineering Ethics Discussion

## Abstract

I distinguish three kinds of engineering: Traditional Engineering of the past, Modernist Engineering of the present, and Focal Engineering of the future. I associate Virtue Ethics (VE) with the *person* who stands out in Traditional Engineering, Conceptual Ethics (CE) with the *process* which stands out in Modernist Engineering, and Material Ethics (ME) with the *product* which stands out in Focal Engineering. All three types of engineering and all three types of ethics are intertwined and cannot be separated, but they can be distinguished. How are these three types of ethics to be adjudicated? I suggest conversations, discourse, the open exchange of information, dialogue, all of which is part of what I call *The Conversation of the Lifeworld.*

Assessments are made within the conversation of the lifeworld. For example, engineers in a group within the workplace can and should dialogue about what makes them morally or ethically good as well as technically good engineers. Being morally good engineers, I suggest, requires the practice of the virtues of honesty, fairness, and care. Opening up a conversation about virtues to be practiced in the workplace requires a point of departure, which can take the form of mathematical Index Functions. The appeal of this to the engineer is obvious. Though ethics is not about numbers, numbers can facilitate entrée into the discussion. Conceptual Ethics, adjudicated at the professional level, and Material Ethics, at the level of the social world, can also benefit from mathematical Index Functions, again, to facilitate entrée into the discussion.

The aim of this paper is to present the mechanism of what I call *The Ethics Engine* involving a set of nested mathematical Index Functions. In addition, I will present a variety of examples showing how a class of approximately 30 students can be broken up into 5 groups of six, each group representing a different company working on part of a new device. Each group can discuss the Virtue Ethics of individual engineers. Then representatives from each group can form a team of professionals who will dialog about the Conceptual Ethics of the companies involved. Then the entire class can join a conversation of the social world within which the engineered product can be assessed from the point of view of Material Ethics.

## Introduction

On one hand, engineers have an affinity for mathematics. A control systems engineer, for example, is constantly manipulating matrices of numbers in various sorts of vector/matrix equations. On the other hand, although the need for engineers to engage with some kind of engineering ethics is growing in the contemporary era, engineers tend to be disinclined toward the concerns of ethics. That does not mean that engineers are unethical. Rather they tend to get caught up in the exigencies of a given engineering project and have little time to spend on the ruminations of ethics. Leave that for the supervisors and managers to worry about. My contention is that if engineering ethics were presented in a mathematical framework it might have more appeal to the typical engineer.

The aim of engineering ethics is to indicate to engineers how they ought to be, what they ought to do, and what they ought to make, in order to bring good into the world. Of course, what we mean by “good” is a wide open question worthy of discussion and reflection. The IEEE Code of Engineering Ethics, as well as many similar codes, gives us an indication of what direction we ought to go. They say, among other things, that to serve the good we ought to protect the health, safety, and welfare of the general public. That is, we should seek to maximize health, safety, and welfare. Mathematically we can indicate this by letting $J_1$ = health, $J_2$ = safety, and $J_3$ = welfare. Then we seek max $J_1$, max $J_2$, and max $J_3$.

In the theory of optimal control, we seek to maximize value functions or, more commonly, to minimize cost functions. But these reduce to the same thing since value functions are usually the negative of cost functions, so that $\min J = \max(-J)$. The control variable that accomplishes this is called the optimal control. In the control problem, the cost function we want to minimize is typically something like energy expenditure or time, whereas in the ethics problem, the value function we want to maximize is typically something like safety or environmental sustainability. Can we design a bridge, for instance, such that the maximum of safety will accrue? Among other things, this is what civil engineers seek to do. They want to maximize safety subject to the various constraints imposed by the problem specifications.

Both the optimal control problem and the problem of engineering ethics have common features. They both entail minimizing a cost (or maximizing a value) subject to a variety of constraints. The optimal control problem, however, can usually be formulated in a rather clear and distinct and objective fashion, while the ethics problem relies on human subjectivity requiring a range of judgment calls. The ethics problem constraints are often not clear and distinct. In the optimal control problem, however, the primary constraint is actually a mathematical model of the system to be controlled, often expressible by a state equation, a differential equation of the form:

$$dx/dt = f(x, u, t)$$

where x is the vector of internal states of the system and u is the vector of control inputs which we are trying to determine. We seek the values of u such that a cost function J is minimized (or a value function is maximized), subject to the constraint of the differential state equation.

## Optimal Control

To sketch out a brief overview of the optimal control problem, it is normal to let the cost function we want to minimize be expressed in the form of an integral:

$$J = \int L(x, u, t)\, dt$$

which, as I mentioned, typically represents something like the energy expenditure in performing the control task. In a realistic problem there are often several constraints that come into play, but I will look at the simplest case where the only constraint is the system state variable equation. To minimize J subject to the constraint of the state equation, we can form what is called a Hamiltonian function:

$$H(x, u, t) = L(x, u, t) + \lambda^T(t)\,[f(x, u, t)]$$

where $\lambda$ is a LaGrange multiplier. The problem of minimization of a cost function subject to a constraint then becomes a minimization of the Hamiltonian without a constraint. Applying the calculus of variations, we get:

$$\partial H/\partial\lambda = dx/dt = f(x, u, t)$$

$$\partial H/\partial x = - d\lambda/dt = [\partial f/\partial x]^T\lambda + \partial L/\partial x$$

$$\partial H/\partial u = 0 = \partial L/\partial u + [\partial f/\partial u]^T\lambda.$$

Solving these equations we get the optimal control which minimizes the cost function J. The resulting equations can be very elegant, especially in the case where the state equation is linear and the cost function is quadratic. This linear-quadratic-optimal control problem requires solution of the famous matrix Riccati equation, named for the 17th/18th century Italian mathematician Jacopo Francesco, Count Riccati.

## Engineering Ethics

In the engineering ethics problem, like in the optimal control problem, the primary constraint would also be a system to be controlled. But in this case the "system" constraining the ethics problem is the scenario within which the ethical issues emerge, as well as the connection of human interactions and ethical deliberations needed to evolve a sound ethical judgment. Such involvements do not lend themselves very easily to mathematical description. Strictly speaking, then, application of optimal control theory to the problem of engineering ethics is not feasible. Furthermore, getting a mathematical description of the cost or value function in the case of the ethics problem does not seem very likely either. Nevertheless, some of the ideas of control theory can be mapped onto the ethics problem.

Even though we lack a clear and distinct and objective cost or value function in the general problem of engineering ethics, and we are required to invoke human judgment calls, ethics does require the notion of values and we can express values numerically. How might one represent things like safe and un-safe numerically? We can propose a scale, say, from minus three to plus three. The minus three would be extremely un-safe and the plus three would be extremely safe. A zero would be half-way in between. As long as everyone involved in the ethical deliberation agreed to such a scale, then the numerical values that emerged would be meaningful.

Even though the world to which the ethics problem is subject has no clear and distinct formulation, we can state the problem in a preliminary fashion as max J, subject to the constraint of the world within which the particular problem unfolds. The engineering ethics problem proceeds in what most people consider a subjective sense. However, I like to see it as not really objectivity or subjectivity but rather as inter-subjectivity wherein via conversations and dialog common ground is laid out. All stake-holders and interested parties have some say in the matter. For example, ten people study the problem, discuss it, assess its ethical content, and decide on values for J: -3, -2.5, -2.0, 0, -2, -1, -3, 0, 1, -2.5, which yields an average of -1.5. On a minus three to plus three scale, this is not a very good number. If honesty, for example, were at issue, we must conclude that dishonesty pretty much prevailed.

Of course, most problems of engineering ethics involve multiple values. I might want to produce products that are safe and at the same time environmentally sustainable. These are both ways to contribute to the good. Again, if I call these values $J_1$ and $J_2$, I would like to have values of plus-three for both. Or if there are three values of interest, I would like to attain the maximum of three value functions, say $J_1$, $J_2$, and $J_3$. The multiple-criteria optimization problem is investigated within the field of Differential Game Theory.

## Differential Game Theory

If the optimal control problem has more than one controller, each of whom seeks a different goal, then the problem is called the differential game problem. Another way to say this is that the differential game with only one player, or one controller, reduces to the optimal control problem. I will look at the three player differential game in which we seek the maximum of three value functions, $J_1$, $J_2$, and $J_3$. The first player chooses $u_1$ such that $J_1$ is maximized, the second player chooses $u_2$ such that $J_2$ is maximized, and the third player chooses $u_3$ such that $J_3$ is maximized. That seems straightforward enough. However, the complications soon become apparent when we realize that each value function is normally dependent on the strategies or controls of all three players. Each player, then, must make some assumptions about the other two players.

Depending on the kinds of assumptions made, different kinds of solution emerge. The three most common types of differential games are the *minimax*, *Nash*, and *Pareto* games. The minimax solution indicates the most competitive and worst case scenario. If I am player one, I seek the max of my value function assuming the other two players are seeking the minimum:

$$\max_{u_1} \min_{u_2,u_3} J_1(u_1, u_2, u_3)$$

And a similar situation exists for the other two players. That is, they seek

$$\max_{u_2} \min_{u_3,u_1} J_2(u_1, u_2, u_3) \qquad \text{and} \qquad \max_{u_3} \min_{u_1,u_2} J_3(u_1, u_2, u_3).$$

The minimax solutions are worst case but they are in a sense the safest. By playing my minimax strategy I may not do so well, but I cannot do any worse, regardless of what the other players do.

Another type of solution that emerges is called the Nash equilibrium solution, named for the famous mathematician John Nash. For this type of game we assume each player is seeking to maximize his or her own value function under the assumption that the other two players are doing the same. The third type of solution, which I stress in the remainder of this paper, is called the Pareto optimal solution, named for the Italian economist, Vilfredo Pareto. It represents the totally cooperative solution and generally yields the best results. However, it is the most dangerous because any one player can deviate from being totally cooperative and improve his or her result (a higher number for the value function can result), and the other two players can end up with diminished results (lower numbers for their value functions). Thus, the Pareto optimal solution entails the risk of "double-cross." But if the Pareto optimal solution can be effected, all players end up with the best possible values for their value functions. Now, for the ethics problem with three values being striven for, I see no reason why absolute cooperation cannot be assumed. All the players in this game are, after all, on the same team.

The way the Pareto optimization problem is set up is to construct a single value function which is a combination of the three separate value functions:

$$J = \alpha_1 J_1 + \alpha_2 J_2 + \alpha_3 J_3$$

where the $\alpha$ terms, called weighting factors, add up to one:

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

and are chosen by discussion and consensus. All the efforts of all three players are then employed in the task of maximizing the value function J. If all three value functions $J_1$, $J_2$, and $J_3$ achieve their maximum value of plus three and they are equally weighted, then each $\alpha$ term is 1/3 and the overall value function J has a value of plus three also. (Actually in this case we would get the plus three result with *any* weighting of the $\alpha$ terms.) This represents the best we can do in any given ethical assessment.

## Three Kinds of Ethics

In previous papers[1,2] I have suggested that there are three different kinds of engineering ethics, which are connected but distinguishable. I called them *virtue ethics*, *conceptual ethics*, and *material ethics*. I also loosely associated virtue ethics with the type of engineering I called Traditional Engineering, which is the engineering of the past. It relied more on intuitions and rules of thumb than does Modernist Engineering, the engineering of contemporary times, which relies more on science and mathematical descriptions of systems to be engineered. With modernist engineering I associated conceptual ethics. Traditional engineering, however, is actually embedded within modernist engineering. Intuitions and rules of thumb are still essential parts of everyday engineering practice. In a similar manner, virtue ethics is subsumed by conceptual ethics in the sense that virtuous engineers practicing, for instance, the virtues of fairness, honesty, and care would be assumed when conceptual ethics is invoked. Then I associated material ethics with a kind of engineering I call Focal Engineering, an engineering of the future. Traditional and modernist engineering are both embedded within focal engineering, because focal engineering employs the tools and techniques and intuitions of traditional and modernist engineering. But focal engineering goes beyond them in aiming for an engineering that *seeks to do good* in the world and contributes to the conversation about the good, rather than just *seeks to do no harm*. The latter, I suggest, is necessary, but only the former is sufficient.

Of the three kinds of ethics, virtue ethics is the ethics of the person. It asks how the person should be in order to be a morally good engineer. Conceptual ethics is the ethics of the process. It asks how the process should be in order to be an ethically good process. Material ethics is the ethics of the product. It asks how the product should be in order to be an ethically good product. Person, process, and product are intertwined, and so are the three types of ethics under consideration. Those connections are manifest in the expression *the engineer engineers the engineered.* Nevertheless, distinctions can be drawn. The three types of ethics can be investigated separately, perhaps as a first cut, the same way that engineers often use linear models for analysis and design work, recognizing the status of those models as approximations to the highly complex and non-linear real world.

I associated with each type of ethics a trilogy of values. With virtue ethics I chose the virtues as values, the virtues of honesty, fairness, and care. With conceptual ethics I chose the values of health & safety, environmental sustainability, and social justice. With material ethics I chose the values of engagement, enlivenment, and resonance. My reasons for choosing these particular nine values, as well as the nature of the values themselves, are discussed in the previously mentioned papers and in a text I am currently writing. Looking at virtue ethics, I can write a value function:

$$J_v = \alpha_v J_{v1} + \beta_v J_{v2} + \gamma_v J_{v3}$$

$$\text{where } \alpha_v + \beta_v + \gamma_v = 1.$$

$J_{v1}$ represents the virtue of honesty. If all engineers involved in the case at hand are being honest with each other, then the value function can be given a plus three. If all are dishonest, the value function can be given a minus three. Similar things can be said for the virtues of fairness and care represented by $J_{v2}$ and $J_{v3}$.

If this problem were a differential game problem, rather than a problem in engineering ethics, then I could express the value functions mathematically, perhaps as integrals representing energy expenditure (or the negative of such integrals) and I could proceed with the maximization of $J_v$ subject to the constraint of the state equation which describes the system to be optimized. Optimization in the Pareto optimal case really reduces quite directly to an optimal control problem. There are Hamiltonian functions involved and a set of partial derivatives, as in the optimal control case. Unfortunately, again, the engineering ethics problem cannot be expressed in such clear and distinct mathematical form. But, also again, we can express value functions numerically and invoke a scale such as minus three to plus three. Then we can make meaningful statements about the best way to assess an engineering ethics scenario.

Now, if I look at conceptual ethics, I can write a value function:

$$J_c = \alpha_c J_{c1} + \beta_c J_{c2} + \gamma_c J_{c3}$$

$$\text{where } \alpha_c + \beta_c + \gamma_c = 1.$$

And in a similar vein I can write a value function for material ethics:

$$J_m = \alpha_m J_{m1} + \beta_m J_{m2} + \gamma_m J_{m3}$$

$$\text{where } \alpha_m + \beta_m + \gamma_m = 1.$$

Once values for each type of ethics are decided upon through negotiation, compromise, and consensus, then an over-all value function can be calculated of the form:

$$J = \alpha J_v + \beta J_c + \gamma J_m$$

$$\text{where} \quad \alpha + \beta + \gamma = 1.$$

The value for J will again range from minus three to plus three. But is such a number an *answer* to our ethical dilemma? No, but it does provide a point of departure from which more rigorous and in depth discussions can unfold. If J came out to be say 2.5 or more we might conclude that there was very little at stake in terms of the ethics of the problem at hand and the engineering involved could be given a green light. But a value lower than say 1.5 could indicate that some serious discussion was in order. In the optimal control problem or the differential game problem, a solution, once obtained, can be implemented by a real piece of physical hardware, for example, like an electrical circuit of some kind. The optimal solution, once implemented, will yield results very much like those predicted by the mathematical equations.

In the engineering ethics problem, on the other hand, the numerical results are more suggestive. They suggest ways to proceed. They help to stimulate the conversation of the lifeworld, which is where real ethical decisions need to be hammered out. That conversation, I submit, should be carried out in three different realms, although those realms are nested. The conversations about virtue ethics should be carried out at the company level, for example, at the level of the engineering team, overseen by a group leader, or perhaps by a company ombudsman. The conversation about conceptual ethics should be carried out at the level of the profession, perhaps by a professional ethics committee. The conversation about material ethics should be carried out at the social level, perhaps employing something like a Danish Consensus Conference Model wherein two panels form, one of experts and one of lay people. They discuss the matter at length and then open the proceedings to the general public. The virtuousness of the engineering team will normally be assumed within the discussions at the level of the profession, and a positive assessment at the professional level will normally be assumed within the conversations at the social level.

The numerical results presented above can inform these conversations and help to orient the discussion. *The Ethics Engine*, which is what I call the mathematical apparatus used to come up with numerical values for J as well as $J_v$, $J_c$ and $J_m$, should be seen as an aid to ethical decision making, not as a methodology for arriving at those decisions. Human judgments cannot be sidestepped in doing engineering ethics. I think this might be a temptation to engineers: if I can solve this engineering ethics problem simply by employing the Ethics Engine, why not do so and then be on with the "more important" tasks of technical engineering? The key thing to stress to engineers and engineering students is that engineering is inherently a social practice. It brings into the world systems, devices, structures, organisms, and networks that have an impact, which may be good, bad, or indifferent. Ethical notions ought to be part and parcel of our everyday engineering work. Ethics should not be just something we look at in our spare time.

## In the Classroom

The Ethics Engine involving a set of nested mathematical Index Functions can be presented in a classroom situation where, for instance, thirty students can be broken up into six groups of five, each group representing a different company, or different team within a company, working on part of a new device. Each group can discuss the virtue ethics of individual engineers. Then representatives from each group can form a team of professionals who dialog about the conceptual ethics of the companies involved. Then the entire class can join a conversation of the lifeworld or social world, within which the engineered product can be assessed from the point of view of material ethics.

Consider the case of engineering a new cell phone. A problem is that the phone requires a certain type of material that is primarily available only in a certain African country run by an oppressive dictator who uses the profits he makes from selling the material to further enslave his own people. The material is also available from another more benign country in South American, but it costs ten times as much. Break the class up into six teams of five students: the overall design group, the group focusing on features A, B, C, the group focusing on features X, Y, Z, the test group, the production group, and the marketing group. Without further background and discussion involving real teams, I will assume the virtuousness of the individual engineers on the engineering teams and will set the overall virtue ethics index at three for all six teams, that is, assume all team players are quite honest, fair, and caring, especially in their interactions with each other. It is when we move to the conceptual ethics and material ethics levels that the problems start.

Select one person from each team, or have the team select a person, to act as a representative to the professional ethics committee which oversees the adjudication of the processes of the engineering enterprise. Here is where conceptual ethics comes into play. Are the values of social justice, health and safety, and environmental sustainability being followed? Obviously the contributions to social justice are being compromised by dealing with the African country and a switch to do business with the South American country should be discussed. Paying ten times more would be prohibitive. Yet there is the company's reputation to consider. It is, after all, *our* company. If word got out that we were supporting a repressive dictator, our image in the public domain would be tarnished. Then there is the idea of "doing the right thing" which would demand that we switch to the South American country. Those discussions continue. But for now, since the company *is* doing business with the African country, we must conclude as a first estimate that the value of the social justice index function should be set to minus three.

Regarding health & safety, there is uncertainty about the electro-magnetic energy that cell phones emit. But since most experts agree that this problem is minimal, we can perhaps give the health & safety index function a value of plus two. Environmental sustainability does not appear to be a major issue. The disposability issue is worthy of some consideration. Recycling possibilities should be discussed. But in general not too many down sides to the environment are apparent, so a value of about plus two would seem to be appropriate here as well.

Now, assume the entire class of thirty engages in a conversation of the lifeworld and embarks on a material ethics assessment. They are assessing the cell phone from the point of view of citizens as well as engineers. What about the engagement, enlivenment, and resonance that the cell phone is supposed to advance? Engagement refers to the harmony between product and end-user. When I pick up a cell phone and use it I engage with another person in an immediate kind of way, but I simultaneously disengage from a wealth of close at hand direct experiences. This is really true of all telephone use. That simultaneity of engagement/disengagement is what needs further and deeper conversation. Is our company's cell phone any different than all the others on the market, as far as this concern is at issue? As a preliminary take on the matter we can give the engagement index a neutral 0.0 value. Enlivenment refers to the harmony between person and her/his world as she/he takes up with the cell phone. Assume the class agrees that conversations on a typical cell phone typically range from the practical to the mundane. Real enlivenment seems to call for face to face involvement. However, thanks to the cell phone, people in distant lands, far from any form of wired communication systems, can now have access to the World Wide Web via cell phone technology and have the world at their finger tips. Via cell phone technology, people in rural Outer Mongolia can connect with their brothers and sisters in downtown Manhattan. Their lives are certainly enlivened. The class decides to give the enlivenment index a positive value of 2.0. Lastly, the resonance issue: resonance refers to the harmony between product and world. This is probably the most difficult feature of material ethics to assess. Certainly the cell phone has transformed the world. We are today a much more connected world. But are we any closer together? As the philosopher Martin Heidegger has said "All distances in time and space are shrinking…yet the frantic abolition of all distances brings no nearness; for nearness does not consist in shortness of distance."[3] So, is the cell phone resonating with the world or is it just another technological device that adds clutter and chatter to an already stuffed and noisy world? Conversations are again required. But in a preliminary fashion, assume the class decides on a neutral value of 0.0 for the resonance index in the material ethics assessment.

Overall, then, with equal weightings on the three values of the material ethics, conceptual ethics, and virtue ethics assessments, we get values of 0.67 for $J_m$, 0.33 for $J_c$, and 3.0 for $J_m$ which implies that there is much more discussion that needs to take place at both the conceptual ethics level and the material ethics level before any kind of ethical judgment can be put forth. Again, the Ethics Engine does not provide solutions to ethical dilemmas, but rather helps to orient us or point us in the right direction for further discussion.

Another scenario might involve a larger class of say 50 students. Five groups of five would represent five different companies. Each group can discuss the virtue ethics of individual engineers. Then a team of professionals could be formed consisting of seven members of a professional ethics committee who will dialog about the conceptual ethics of the companies involved. Then the entire class can join a conversation of the social world, via a Danish Consensus Conference Model engagement, within which the engineered product can be assessed from the point of view of material ethics. Two panels would be formed, one of experts about the technology at issue and the other of lay people. Each panel would have nine members. After the experts inform the lay panel, the discussion is opened to all interested parties. That would include the whole class minus the two panels: 50-18 = 32 people. The details of a conversation of the lifeworld based on a Danish Consensus Conference Model engagement would be discussed ahead of time.

Other classroom discussions can be imagined. Such scenarios would give engineering students some sense of the social implications of the work that they do. In particular, the engineered products they bring into the world ought to be looked at carefully in order to ascertain whether or not these products will contribute to "the good." What we or they mean by the good is another other conversation which must remain on-going.

## Bibliography

1 Gene Moriarty, “Three Kinds of Ethics for Three kinds of Engineering,” *IEEE Technology and Society Magazine*, Fall 2001.

2 Gene Moriarty, “What, More Assessment?," ASEE Conference (June 2001) in Albuquerque, NM.

3 Martin Heidegger, “The Thing,” an essay in the book *Poetry, Language, Thought*, translated by Albert Hofstadter. Harper & Row: New York, 1971, p. 165.