Imagine there's no countries...and no religion too.

…I hope someday you'll join us, and the world will be as one.

John Lennon (*Imagine*)

Chapter 3: **Natural Gadfly**

One of my first jobs out of college was to serve as research assistant to Carol Krumhansl, a pioneer in the field of psychology of music. Before I joined her team, Carol recorded a world-class pianist's rendition of a Mozart piano sonata using a robotic piano. It was equipped with force-sensors under each key so that it could precisely replicate the performance of anyone who played it. The pianist deviated from Mozart's sheet music only dynamically and rhythmically, so I had the piano create three more versions: correcting the dynamic deviations, the rhythmic deviations, and both. The remainder of my job was to welcome college students into the lab, have them listen to each of the four versions in random order, and record their ratings of "tension in the music" as they listened.

Carol considered the experiment a failure. The average measure-by-measure tension ratings for all four versions were equivalent. There was no measurable evidence that the pianist had improved upon Mozart's sheet music nor that Mozart relies on anyone to "interpret" his music. The pianist's deviations seemed to be just an artifact of performance uniqueness—we might expect the same pianist to make different departures each time. From the standpoint of enduring musical truth, the world-class pianist apparently contributed nothing.

That raised the question of why the four versions differed at all. Were the pianist's aberrations errors? Was the pianist unable to perform the sonata as written? Are humans

simply inferior to machines when it comes to performing music? It may be difficult for humans to perform as consistently as machines, but that doesn't necessarily make us inferior musicians. Maybe world-class pianists deviate from sheet music and create unique performances on purpose. Maybe they nurture a habit for creativity, exploring and experimenting when they perform, allowing unforeseeable feelings to influence them on the off-chance it might impact a listener in an unpredictably better way.

I did not tell listeners anything about the differences between the four versions of the piano sonata, nor did I ask whether they could distinguish them, but they inevitably volunteered that they detected distinctions, and identified the strictly mechanical performance as the worst. It was as though they thought I was testing their taste in classical music, and wanted me to confirm that they passed. Apparently they thought there is something good about a pianist being creative, deviant, inconsistent, and attempting to stir audiences in unforeseeable ways.

If you are a natural gadfly, then you prefer to be creative, deviant, inconsistent, and provocative even beyond the context of the arts. Like improvising comedy, you experiment with what might turn-out better than anyone could have expected. Making new discoveries that could persist long after you pass away, you aim at a perfection you expect no one, including yourself, to fully understand in your own lifetime. To defend against immoral impulses, you may "pick your battles" like a pianist who mostly follows the sheet music, but you prioritize social progress and escape from bias.

If you are not a natural gadfly, then learning to appreciate these artists may be as difficult as learning a new language, or as understanding what it is like to be born blind. This

chapter cannot neutralize that challenge, but it will hopefully convince you that it is worthwhile.

Your Heritage as a Natural Gadfly

Natural gadflies inherit a wealth of theories, stories, laws, traditions, organizations, and technologies tuned to meet their needs.

The term "social gadfly" is typically traced to Plato's account of the trial of Socrates in ancient Athens. Socrates is considered one of the fathers of western philosophy, and one of the most respected people of all time. He was put on trial around the age of seventy because he made a habit of striking-up conversations in the public marketplace like this: “You are much wiser than me. Please, I beg you, enlighten me how is it that you reconcile this apparent contradiction in your morals…” after which he would point-out some flaw embarrassingly beyond explanation. Like a modern journalist criticizing celebrities and politicians, Socrates was perceived as a threat to authorities, and some of his potential victims retaliated through the courts.

A jury of about 500 ancient Athenian citizens convened to hear arguments and vote on whether Socrates' behavior should be allowed. According to Plato, Socrates spoke eloquently in his own defense, but in ways he knew would insult his jury. His argument culminated, "Athenians, I am not going to argue for my own sake…but for yours…For if you kill me you will not easily find a successor to me, who, if I may use such a ludicrous figure of speech, am a sort of gadfly, given to the state by God; and the state is a great and noble steed who is tardy in his motions owing to his very size, and requires to be stirred into life. I am that gadfly which

God has attached to the state, and all day long and in all places am always fastening upon you, arousing and persuading and reproaching you. You will not easily find another like me, and therefore I would advise you to spare me."

To Socrates' surprise, he was nearly declared innocent, only about thirty votes shy of the two-hundred fifty votes he needed. A second trial was held to determine his punishment. In the second trial, he was sentenced to death, although many jurors may have expected him to flee his execution (i.e. they may have intended to segregate, rather than punish). Citing his loyalty to the law, Socrates stayed in Athens, and carried out his sentence by freely drinking a poison made of hemlock.

Respected social gadflies both preceded and succeeded Socrates—just about every major social reform, institution or theory is credited to one gadfly or another—but Socrates is worth special mention in this chapter because you inherit several important tools from his martyrdom: First of all, his death inspired his students to create a safer place for gadflies. It was called "the Academy." Academia has grown into a worldwide community of schools and laboratories. Its focus has shifted from protection of unpopular questions to education of our workforce, but it nonetheless continues to feature resources which nurture gadflies like you.

Another tool you inherit from Socrates' martyrdom is freedom of speech, a legal innovation, dating back to the Roman Empire, which recently bloomed into an international human right. It is aimed at protecting others from Socrates' fate. It doesn't protect hate speech, pornography or offensive speech, malicious lies, invasions of privacy, copyright violation, or slander—it only protects speech which might enable social progress. Thus, it functions as a legal mandate for gadfly behavior, and has served as a catalyst for other legal innovations to

support you: freedom of expression, freedom of the press, freedom of thought, scientific freedom, and even freedom of information.

Other gadfly heirlooms have less-ancient origins. Since the industrial revolution, various philosophers, especially John Dewey, have re-articulated Socrates' gadfly theory in terms of science and social progress. This re-articulation, often called "pragmatic ethics," holds that moral truth, like scientific truth, is discovered by society over the course of many generations. For practical reasons, scientists must usually act as though their current theories are perfectly true, yet they also engage in experiments which aim to supplant those theories. Similarly, pragmatic ethicists say we should embrace current moral norms, yet engage in what John Stuart Mill called "experiments of living" which have the potential to overturns those norms.

Alfred Nobel, a contemporary of the pragmatists, established a set of prizes which have been awarded annually for over a century to recognize social reformers as well as scientists. Through such forms of recognition, including intellectual property rights and grants, gadflies have gained a position of respect. At first, that position was reserved for gadflies lucky enough to "succeed," but the emerging academic fields of innovation, entrepreneurship, social innovation, and social entrepreneurship (i.e. studies of social change) now tell us that we should honor "failed" reforms as well because social progress rarely comes without failure. Thus, at least in these circles, gadflies can find respect and support even before fame (e.g. see Ashoka, the Skoll Foundation, the Omidyar Network, the Schwab Foundation, Athgo, Root Cause, the Canadian Social Entrepreneurship Foundation, NESsT, New Profit Inc., National Social Entrepreneurship Forum, and Echoing Green).

Today, the largest treasures being added to your inheritance are technologies. In many ways, the World Wide Web, perhaps the greatest technological marvel in existence, is designed for gadflies. Search engines, Wikipedia, online petitions, anonymizing newsleak sites, TED talks, YouTube, and social networking are all designed to provide channels through which gadflies can change the world. For example, Justin Bieber and the founders of Facebook both launched their movements through the World Wide Web. If a gadfly wanted to gather survey data to refute the claims of this book, they could likewise gather it inexpensively through Amazon's online "Mechanical Turk." Compared to Socrates, modern social gadflies inherit a wealth of support and protection.

The Social Importance of Gadflies

There are two ways to explain the social importance of gadflies. One way is to point to evidence in psychology, sociology, economics and history. We will return to that approach in a few pages, but first we will discuss evidence from computer science.

Studying computers to learn about human minds is a bit like studying cells to learn about bodies; it can be a valuable research technique, even though it reveals only part of the story. Modern computers do not share the human capacity to be GRIN-closeted, nor do they share key human experiences—being loved, being independent, and participating in institutions (such as families)—but computers do exist of each GRIN-type, and evaluations of those computers can help us understand GRIN dynamics among humans.

Any software which uses random numbers to behave in ways its programmer cannot predict has gadfly orientation. Software is not human. It lacks human knowledge. But the

behavior of such software is otherwise like the behavior of humans exhibiting gadfly orientation. Such algorithms form the backbone of "evolutionary computation," a form of decision-making typically involving populations of virtual machines, each of which makes unpredictable deviations from the status quo. One or more non-gadfly machines evaluate the creations resulting from these deviations, and the creations deemed most successful set the new status quo from which the next generation of deviations is made.

Evolutionary computation is inefficient in many situations, but no alternative explores as wide a range of possibilities, so gadfly machines are the best for automated discovery and invention. Thus far, they have invented better engines, turbines, microchips, antennae, optics, drugs, cleansers, scientific theories, and strategies for marketing, financial trading, image processing, game play, and military contests. They have solved tough mathematical problems, created art, and composed music performed by the London Symphony Orchestra, and even jammed with human musicians live. To put it bluntly, these machines have filled the social role Socrates described for gadflies, and giving credit for that to their programmers would be like crediting Martin Luther King Jr.'s *parents* for his "I Have a Dream" speech; mature gadfly machines are just as unpredictable as grown children.

In technical jargon, gadfly machines outperform non-gadfly machines on "problem spaces with local maxima." This phrase derives from the hypothetical situation of a blind man trying to reach the highest peak in an uncharted territory, what would be called its "global maximum." The blind man cannot see distant peaks, but he can measure his current altitude and recall the altitudes of all the places he has been, so he can tell which direction goes "uphill," and keep climbing whatever hill he happens to be on until he reaches its peak. When he stands atop the highest hill he has encountered thus far, however, he cannot tell whether higher peaks

exist elsewhere in the territory. He knows that he is at a maximum, but not whether that maximum is "global" or merely "local." Non-gadfly machines in that situation get stuck on that peak even if it isn't global, but gadfly machines engage in what is frequently called "mutation" or "random restart," which allows them to explore elsewhere.

Abandoning the highest known peak is a waste of time in territories which have been fully charted already or which have only one peak. Furthermore, if the territory is small enough, it would be more efficient to explore it systematically. Thus, gadflies have advantage only in a certain kind of territory, in situations in which it is difficult to recognize when a strategy is suboptimal. That is true in science, where thousands of geniuses fail to recognize flaws in current theory. It is also true in invention and fashion, where thousands of geniuses fail to anticipate what will replace the current state of the art. Certain aspects of the history of morality, such as changing attitudes about racism, gender, torture, polygamy, idolatry, monarchy, and the environment suggest that the quest for morality is plagued by similar local maxima, and would therefore benefit from the contribution of gadflies.

Although gadflies are essential components of evolutionary computation, they are subject in such systems to negotiators who restrict the information available to them, limit the territories they can explore, and determine when they will start over. If those gadflies were human, such restrictions might qualify as violations of the right to freedom of thought. That raises the question of whether evolutionary computation could be improved by extending such freedom to machines.

Giving gadfly machines greater freedom, it turns out, can greatly improve results. As an example, typical application of evolutionary computation to circuit design involves judging

designs based on their theoretical merits, but Adrian Thompson's implementation involved testing actual renderings of the designs. In other words, the gadflies had their blinders relaxed; they were allowed to tinker with real hardware. After reinitializing his gadflies about three-thousand times to the best designs discovered so far, Thompson had a working prototype which required only thirty-seven components. Returning to the drawing board so many times may sound tedious, but this end result was ten times more efficient than the best design produced by any human ever. It worked by exploiting electromagnetic phenomena beyond current circuit design theory. In other words, Thompson's implementation granted gadfly machines the freedom to reject theories he himself took for granted, and that freedom was necessary to achieve the unprecedented level of success his gadflies did.

In retrospect, this is not surprising. Unless inherited theories are already perfect, the path to perfection must require the freedom to discard them. Yet the potential for unpredictable results can be frightening, so our policy for dealing with machines today is like that for dealing with five-year-old children; we have no intention of trusting them with complete freedom just yet. Much as ancient Athenians considered an unrestrained Socrates dangerous, we now label most unrestrained software, "viruses." If machines become more trusted in the future, computer science might provide all the evidence needed to substantiate the social importance of gadflies, but we must supplement with evidence from the social sciences in the meanwhile.

In July 1961, responding to Nazi war-crime trials, Yale University psychologist Stanley Milgram conducted what may be the most famous social psychology experiment ever. Two people, a subject and a confederate, come to the lab. They draw slips of paper to divvy the roles of "Teacher" and "Learner." The subject always gets the Teacher role. The confederate (Learner)

is strapped into an electric chair, then an Experimenter brings the Teacher to an adjacent room which has 30 buttons for delivering shocks ranging from 15 to 450 volts. The Experimenter instructs the Teacher to teach word pairs to the Learner via intercom, and to administer a shock for each wrong answer, increasing the voltage each time. Rather than deliver shocks, the buttons actually trigger recorded sounds which make it appear as though the Learner is being shocked, but Teachers really do believe they are hurting the Learner. The purpose of the experiment is to measure how many people would press the 450-volt button, which is clearly labeled, "Lethal."

The Milgram Experiment has been repeated by many psychologists across cultures and genders. All found that 61-66% of us would deliver the lethal shock. Milgram concluded that "relatively few people have the resources needed to resist authority," but escaping the situation he engineered not only requires Teachers to disobey the authority of the Experimenter—it also requires them to relinquish their teaching goal, and to prioritize a distant relationship over a closer one (Milgram found that significantly more subjects pass the test when the Learner is less remote). Furthermore, Milgram put his subjects in a situation for which they lacked expertise or previous experience. Being a subject in the Milgram Experiment is very much like being placed on a hill in uncharted territory, such that one would have to sacrifice moral altitude in order to explore the possibility of higher moral ground.

One might classify the Milgram experiment with other attempts at moral manipulation. Gadflies are difficult to manipulate because they are unpredictable. Being unpredictable means that poorly-engineered cons will work surprisingly well on gadflies, but it also means that well-engineered cons will work surprisingly poorly on them. Thus, the more effective the manipulation, the more likely it is that only gadflies will escape. As far as I know, psychologists

have yet to test the hypothesis that people who pass Milgram's test tend to be natural gadflies. Given the nastiness of such testing, however, it might be best to simply recognize that this kind of research would eventually produce a trap which only gadflies can escape.

Milgram's experiment is associated with the Holocaust, implying that the massacre of millions of Jews was made possible through social engineering and that natural gadflies are our best defense against such engineering. Yet, it would be a mistake to expect to encounter local maxima only in engineered situations. Scientists, inventors, fashion designers, and social reformers all face local maxima naturally. By abandoning such maxima, gadflies have innovated better vehicles, better communication devices, better home appliances, and better entertainment devices which bring us comfort, convenience, and efficiency. Joseph Schumpeter's economic theory that innovation is the key to economic success has been confirmed in the Chinese, Indian, and U.S. economies recently, where support for gadfly behavior has been shown to correlate with growth of wealth and GDP. The social importance of gadflies in these cases derives less from moral traps than from the fact that life simply demands progress.

None of this evidence indicates that gadflies have the *only* GRIN orientation a society needs, but it does strongly suggest that a society without gadflies would fall prey to moral corruption, would fail in competition against other societies, and would lack the comforts, convenience, and efficiencies associated with advanced technology and social progress.

What Natural Gadflies Need

How can a society support gadflies, and enjoy the benefits they bring? The neurochemistry of gadfly thought is less well understood than that involved in other GRIN orientations, but clearly involves creativity. A gadfly who cannot be creative herself could solicit creativity from others through whistle-blowing, martyrdom or terrorism, but this would do no good in environments which suppress creativity across the board.

In 2010 Jennifer Mueller from the Wharton Business School published a study of bias against creativity. She found that people claim to favor creativity over practicality, but can be shown to be biased in the opposite direction by measuring the relative time it takes them to associate the two approaches with positive images. In a recent worldwide poll, CEOs reported that creativity is the single most important leadership trait for success, but Mueller demonstrated that we don't back-up such theories in practice. Like racists who condemn racism but remain racist deep-down inside, we subconsciously discriminate against the very thing we praise.

Creativity requires focused effort, access to truth, and audience. Societies disable gadflies by suppressing these essential nutrients. Focused effort can be suppressed through distraction. You cannot be creative when you are busy with non-creative tasks. For example, victims of the Holocaust were distracted with forced labor. As another example, an employer might reduce her workforce to increase efficiency, thus assigning larger numbers of tasks to each remaining employee and diminishing opportunity to focus. Alternately, success as a gadfly might bring a flood of fan mail which takes so much time to review that the gadfly can no longer focus on being creative. Even for gadflies who are not famous, the invention and

development of compelling distractions, such as TV and video games, may exhaust free time that would otherwise be spent creatively.

Effective creativity also requires access to truth. For example, the Holocaust was kept secret in many ways, and effective responses to it emerged only as truth came out. Yet environments need not be secretive or deceptive to block access to truth—sometimes truth is so expensive that environments can block creativity simply by neglecting to supply special resources. Physicists need supercolliders, and artists need inspiring experiences. Without access to truth, gadflies cannot bring benefit to society. A creative process built on lies, or lacking data, merely yields more of the same.

Finally, quality of creativity depends upon quality of audience—at least that's the way it works with creative machines. The effectiveness of their creativity relies on the ability of their audience to filter through their experiments. Creative genius may include skill at being one's own filter, but no genius can replace the filtering-power of a diverse society which takes gadflies seriously, tries their products, and tests their proposals. That may explain why Socrates approached other Athenians in the market, and why he refused to live without access to them.

If you are a natural gadfly, you might assume everyone wants to be creative, but many people seek gadfly roles only to misappropriate gadfly resources. The potential to steal from society by pretending to be creative is a serious problem. Improved control over the use of gadfly resources may help explain why social progress has accelerated across history—it took centuries to develop systems of taxation and specialization that could bathe some gadflies in the nutrients required for greater innovation (e.g. in think-tanks, R&D departments, concert halls,

etc.). Yet, even after centuries of development, these systems remain controversial. They do not support all gadflies, nor do they support any gadfly all the time. Effectiveness as a gadfly depends on social support, and that is far from guaranteed in modern society.

For this reason, your needs as a gadfly also include tolerance. Imagine a gadfly's first performance review: In an ideal situation, the manager might say, "We've noticed that you tend to innovate, and that your innovations aim to transform our organization. Every organization needs to progress—and the value of each hour of progress compounds across our organization's entire life—so we really appreciate your contribution, and have analyzed and tested your innovations to help you perfect them. Innovation is a process of trial and error. We are documenting everything your initiatives have taught us, and sharing that learning across our organization. Meanwhile, we want you to keep going. Here is a list of the specific problems we hope you might find inspiring..."

Realistically, however, after praising gadflies for their initiative and passion for quality, compassionate managers typically say something like this, "There is also an area in which I would like for you to improve: Success requires teamwork and trust. I don't think you mean it this way, but several people have mentioned to me, and I myself have noticed, that sometimes the way you say things gives people the impression that you think we're all incompetent. We do good work, and changing course would be unreasonably risky. I think you know that, but sometimes the way you express yourself creates the impression that you'd like to change the way things are done around here. You don't want to create that impression, do you? I bring this up only because I want you to be successful. I want you to be trusted by your coworkers."

In other words, what often blocks natural gadflies from contributing effectively is intolerance. In the service of progress, you may be willing to question competence, take risks, and expect people to get outside their comfort zones—progress can't happen any other way—but non-gadflies do not share this willingness, and do not want you to subject them to such ambiguity, risk, and discomfort. They may care about progress theoretically, but exhibit other priorities in practice. Entrepreneurship scholars cite this conflict to explain why attempts to reform from within almost never work, why successful change has almost always involved new leaders. GRIN diversity needs as much protection as any other form of diversity, and the message communicated by the manager in the second performance review is that protection will be denied, that being a gadfly in this workplace will be as thankless and difficult as being a stinging fly trying to goad a lazy horse.

To be tolerated, each orientation needs special license, and gadflies need special license to doubt inherited values. If such license were exercised by everyone, that inheritance would disintegrate. That's why the license must be special. Are gadflies smarter than the entire history of geniuses from whom we inherit our current traditions? Probably not, but a society that does not indulge such arrogant fantasies cannot move forward. Even if gadflies are less intelligent, they are the only ones inclined to go where progress demands, so, as unjustified and unfair as it might seem, they need to be treated as the next Socrates, the next da Vinci, the next Moses, the next Steve Jobs.

On the other hand, gadflies also need to learn to recognize the importance of other orientations, to accept a balance in which they contribute, but do not rule. Part Two of this book describes teachings which balance the GRIN-orientations. Taken as a whole, this set of

teachings shows how each orientation needs the others. Gadflies should take special note of Chapter Fifteen, "Deviating from Deviance."