Principles of Effective Research

By Michael A. Nielsen

July 2004

Overview

This essay is intended as a letter to both myself and others, to hold up in the sharpest possible terms an ideal of research I believe is worth working toward. I've deliberately limited the essay to 10 pages, hoping that the resulting omissions are compensated by the forced brevity. This is a rather personal essay; it's not the sort of thing I'd usually make publicly available. I've made the essay public in order to heighten my commitment to the project, and in the hope that other people will find it stimulating, and perhaps offer some thoughts of their own.

A few words of warning. My primary audience is myself, and some of the advice is specific to my career situation [*], and therefore may not be directly applicable to others. And, of course, it's all just my opinion anyway. I hope, however, that it'll still be stimulating and helpful.

[*] I'm a theoretical physicist; I lead a small research group at a large Australian University; I have a permanent position, with no teaching duties for the next few years; I have several colleagues on the faculty with closely related interests.

The philosophy underlying the essay is based on a famous quote attributed to Aristotle: *"We are what we repeatedly do. Excellence, then, is not an act but a habit."* Underlying all our habits are models (often unconscious) of how the world works. I'm writing this essay to develop an improved personal model of how to be an effective researcher, a model that can be used as the basis for concrete actions leading to the development of new habits.

Fundamental principles

The fundamental principles of effective research are extremely similar to those for effectiveness in any other part of life. Although the principles are common sense, that doesn't mean they're common practice, nor does it mean that they're easy to internalize. Personally, I find it a constant battle to act in accord with these principles, a battle requiring ongoing reflection, rediscovery and renewed commitment.

Integrating research into the rest of your life

Research is, of course, only a part of life, and must be understood in relation to the rest of life. The foundation of effective research is a strong *motivation* or *desire* to do research. If research is not incredibly exciting, rewarding and enjoyable, at least some of the time, then why not do something else that is? For the purposes of this essay, I'll assume that you already have a strong desire to do research [*].

[*] People sometimes act or talk as though desire and motivation cannot be changed. Within limits, I think that's wrong, and we can mold our own motivations. But that's a subject for another essay.

Motivation and desire alone are not enough. You also need to have the rest of your life in order to be an effective researcher. Make sure you're fit. Look after your health. Spend high quality time with your family. Have fun. These things require a *lot* of thought and effort to get right. If you don't get them right, not only will your life as a whole be less good, your research will suffer. So get these things right, and make sure they're integrated with your research life.

As an example, I once spent three years co-authoring a technical book, and for the final eighteen months I concentrated on the book almost exclusively, to the neglect of my health, relationships, and other research. It is tempting to ask the question "Was the neglect worth the benefits?" But that is the wrong question, for while the neglect paid short-term dividends in increased productivity, over the total period of writing the book I believe it probably *cost* me productivity, and it certainly did after the book was complete. So not only did I become less fit and healthy, and see my relationships suffer, the book took longer to complete than if I'd had my life in better order.

Principles of personal behaviour: proactivity, vision, and discipline

I believe that the foundation of effective research is to internalize a *strong vision* of what you want to achieve, to work *proactively* towards that vision, taking *personal responsibility* for successes and failures. You need to develop *disciplined* work habits, and to achieve balance between *self-development* and the actual *creative research process*.

Proactivity and personal responsibility

Effective people are proactive and take personal responsibility for the events in their lives. They form a vision of how they want their life to be, and work toward achieving that vision. They identify problems in their lives, and work toward solutions to those problems.

Isn't this obvious, banal advice? I heard a story years ago in which a representative from McDonald's was asked what gave McDonald's the edge in the fast food industry. They replied that McDonald's took care of the little things, like making sure that their restaurants and surrounds were always extremely clean. Representatives of other fast food companies replied incredulously that surely that was not the reason McDonald's did so well, for "anyone could do that". "But only McDonald's does" was the response. The heart of personal effectiveness is not necessarily any special knowledge or secret: it is doing the basics consistently well.

When it comes to proactivity and responsibility, it seems to be incredibly difficult to internalize these principles and act on them consistently. Almost everyone says and thinks they are proactive and responsible, but how many of us truly respond to the force of external circumstance in the most proactive manner?

My belief is that the reason it is difficult to be consistently proactive and responsible is that over the short term it is often significantly *easier* to abdicate responsibility and behave in a reactive fashion. In my opinion, there are three basic ways this can occur.

The first way is to blame external circumstances for our problems. “We don’t have enough grant money.” “I have to teach too much.” “My supervisor is no good.” “My students are no good.” “I don’t have enough time for research.” When challenged on what actions we are taking to rectify the situation, we will claim that it’s the fault of other people, or of circumstances beyond our control, relieving ourselves of the burden of doing anything to solve the problem.

In short, we abdicate responsibility, preferring to blame others. This is easier over the short term, since it’s easier to complain than it is to take action, but is not a recipe for long-term happiness or effectiveness. Furthermore, we will usually *deny* that it is within our power to take actions to improve our situation. After all, if it was in our power, it would be *us* who is responsible, and our entire worldview is based upon blaming others for our own problems.

The second way of abdicating responsibility is to get caught up in displacement activities. These may give us a short-term fix, especially if they win us the approbation of other people, perhaps for responding to requests that they label urgent. Over the long run such displacement activities are ultimately unfulfilling, representing time lost from our lives.

The third way of abdicating responsibility is by getting down on yourself, worrying and feeling bad for not overcoming one’s difficulties. Winston Churchill spoke of the “black dog” of depression that overtook him during times when his political career was in eclipse. Personally, I sometimes get really down when things are not going well, and get caught up in a cycle of worry and analysis, without constructively addressing my problems. Of course, the right way to respond to a bad situation is not to beat yourself up, but rather to admit that, yes, things are going badly, to figure out exactly what problems you are facing, write out possible solutions, prioritize and implement them, without getting too worried or hamstrung by the whole process.

Why are these three options so attractive? Why do we so often choose to respond in this way to the challenges of life rather than taking things on with a proactive attitude that acknowledges that we’re responsible for our own life? What all three options share in common is that over the short-term abdicating responsibility for our problems is easier than taking responsibility for meeting the challenges of life.

A specific example that I believe speaks to many of us is when we’re having some sort of difficulty or conflict with another person. How many of us put off confronting the problem, preferring instead to hope that the problem will resolve itself? Yet, properly managed – a difficult thing to do, most likely requiring considerable preparation and aforethought – it's nearly always better to talk with the person about the problem until you arrive at a mutual understanding of *both* your points of view, *both* sets of interests, and can resolve the issue on a basis of shared trust.

How can we learn to become proactive? I don't know of any easy way. One powerful way is to be inspired by examples of proactive people. This can either be through direct personal contact, or indirectly through biographies, history, movies and so on. I like to set aside regular time for such activities. Another powerful tool for learning proactivity is to remind ourselves regularly of the costs and benefits of proactivity and responsibility versus reactivity and irresponsibility. These costs and benefits are easy to forget, unless you're constantly being reminded that complaints, self-doubt, blame of others and of self are actually the *easy* short-term way out, and that chances are that you can construct a better life for yourself, at the cost of needing to do some hard work over the short term.

In the context of research, this means constantly reminding yourself that you are the person ultimately responsible for your research effectiveness. Not the institution you find yourself in. Not your colleagues, or supervisor. Not the society you are living in. All these things *influence* your research career, and may be either a help or a hindrance (more on that later), but in the final analysis if things are not working well it is up to you to take charge and change them.

Vision

Effective people have a vision of what they'd like to achieve. Ideally, such a vision incorporates both long-term values and goals, as well as shorter-term goals. A good vision answers questions like: What sort of researcher would I like to become? What areas of research am I interested in? How am I going to achieve competence in those areas? Why are those areas interesting? How am I going to continue growing and expanding my horizons? What short-term steps will I take to achieve those goals? How will I balance the long-term goals with the short-term realities of the situation I find myself in? For example, if you're in a temporary job and need to get another job soon, it's probably not such a great idea to devote all your time to learning some new subject, without any visible outcome.

A vision is not something you develop overnight. You need to work at it, putting time aside for the process, and learning to integrate it into your everyday life. It's a challenging process, but over the long run it's also extremely rewarding. History shows that great actions usually are the outcome of great purpose, even if the action that resulted was not the original purpose. Your vision doesn't always need to be of a great purpose; it's good to work on the little stuff, some of the time. But you should occasionally set yourself some big, ambitious goal, a goal that gets you excited, that makes you want to get up in the morning, and where you've *developed* a confidence in your own mind that you have a chance of achieving that goal. Such a great purpose inspires in a way that the humdrum cannot; it makes things exciting and worthwhile if you feel you're working towards some genuinely worthy end. I believe this is particularly important in the more abstract parts of research (like theoretical physics), where it can require some work to make a personal, emotional connection to one's own research. Having a clear vision of a great end is one very good way of making such a connection. When you don't do this, you can get stuck in the rut of the everyday; you need to get out of that rut, to develop a bigger vision.

Finally, a good vision is not inflexible. It's something that gets changed as you go along, never lightly, but frequently. The importance of having the vision is that it informs your everyday and every week decisions, giving you a genuinely exciting goal to work towards.

Self-discipline

Effective people are *self-disciplined*. They work both hard and smart, in the belief that you reap what you sow. How does one achieve such self-discipline? It's a difficult problem. Wayne Bennett, one of the most successful coaches in the history of the sport of Rugby League, sums the problem up well when he says "I've had more trouble with myself than any other man I've ever met".

It is a tempting but ultimately counterproductive fallacy to believe that self-discipline is merely a matter of *will*, of deciding what it is that you want to do, and then doing it. Many other factors affect self-discipline, and it's important to understand those other factors. Furthermore, if you believe that it's all a matter of willpower then you're likely to get rather depressed when you fall short, sapping your confidenc, and resulting in less disciplined behaviour.

I now describe three factors important in achieving self-discipline.

The first factor is having *clarity* about *what* one wants to achieve, *why* one wants to achieve it, and *how* to go about achieving it. It's easy to work hard if you're clear about these three things, and you're excited about what you're doing. Conversely, I think the main cause of aimlessness and procrastination is when you lack clarity on one or more of these points.

The second factor affecting self-discipline is one's social environment. Researchers are typically under little immediate social pressure to produce research results. Contrast this with the example of professional athletes, who often have an entire support system of coaches, managers and trainers in place, focused around the task of increasing their effectiveness. When a researcher stays out late, sleeps in, and gets a late start, no-one minds; when a professional athlete does, they're likely to receive a blast from their coach.

Access to a social environment which encourages and supports the development of research skills and research excellence can make an enormous difference to all aspect of one's research, including self-discipline. The key is to be *accountable* to other people. Some simple ways of achieving such accountability are to take on students, to collaborate with colleagues, or to set up mentoring relationships with colleagues.

The third factor affecting self-discipline is a special kind of honesty, honesty to oneself, about oneself. It's extremely easy to kid ourselves about what we do and who we are. A colleague once told me of a friend of his who for some time used a stopwatch to keep track of how much research work he did each week. He was shocked to discover that after factoring in all the other activities he engaged in each day – interruptions, email, surfing the net, the phone, fruitless meetings, chatting with friends, and so on – he was averaging only half an hour of research per day. I wouldn't be surprised if this was typical of many researchers. The good news, of course, is that building this kind of awareness lays the foundation for personal change, for achieving congruence between our *behavioural goals* and how we *actually* behave, in short, for achieving self-discipline.

Aspects of research: self-development and the creative process

Research involves two main aspects, *self-development* and the *creative process* of research. We'll discuss the specifics of each aspect below, but for now I want to concentrate on the problem of achieving *balance* between the two, for I believe it is a common and significant mistake to concentrate too much on one aspect to the exclusion of the other.

People who concentrate mostly on self-development usually make early exits from their research careers. They may be brilliant and knowledgeable, but they fail to realize their responsibility to make a contribution to the wider community. The academic system usually ensures that this failure is recognized, and they consequently have great difficulty getting jobs. Although this is an important problem, in this essay I will focus mostly on the converse problem, the problem of focusing too much on creative research, to the exclusion of self-development.

There are a lot of incentives for people to concentrate on creative research to the exclusion of self-development. Throughout one's research career, but particularly early on, there are many advantages to publishing lots of papers. Within limits, this is a good thing, especially for young researchers: it brings you into the community of researchers; it gives you the opportunity to learn how to write well, and give good presentations; it can help keep you motivated. I believe all researchers should publish at least a few papers each year, essentially as an obligation to the research and wider community; they should make some contribution, even if only a small one, on a relatively unimportant topic.

However, some people end up obsessed with writing as many papers as possible, as quickly as possible. While the short-term rewards of this are attractive (jobs, grants, reputation and prizes), the long-term costs are significant. In particular, it can lead to stagnation, and plateauing as a researcher. To achieve one's full potential requires a balancing act: making a significant and regular enough research contribution to enable oneself to get and keep good jobs, while continuing to develop one's talents, constantly renewing and replenishing oneself. In particular, once one has achieved a certain amount of job security (a long-term or permanent job) it may make sense to shift the balance so that self-development takes on a larger role.

For many people (myself included) who have concentrated mainly on making creative research contributions earlier in their careers, this can be a difficult adjustment to make, as it requires changing one's sense of what is important. Furthermore, there is a constant pull towards concentrating on research over self-development, since there are often short-term incentives to sacrifice self-development for research ("I've got to get this paper out now"), but rarely vice versa. To balance these tendencies, we need to remember that nobody, no matter how talented, is born an effective researcher; that distinction can only be obtained after a considerable amount of hard work and personal change, and there is no reason to suppose that just because one is now able to publish lots of papers that one has peaked as a researcher.

In my opinion, creative research is best viewed as an *extension* of self-development, especially an extension of a well-developed reading program. I don't believe the two can be completely pried apart, as the two interact in interesting non-linear ways. I'm now going to talk in a little more detail about both processes, keeping in mind that the ultimate goal of research is new ideas, insights, tools and technologies, and this goal must inform the process of self-development.

Developing research strengths

The foundation is a plan for the development of research strengths. What are you interested in? Given your interests, what are you going to try to learn? The plan needs to be driven by your research goals, but should balance short-term and long-term considerations. Some time should be spent on things that appear very likely to lead to short-term research payoff. Equally well, some time needs to be allocated to the development of strengths that may not have much immediate pay-off, but over the longer-term will have a considerable payoff.

In targeting areas of development, an important goal to keep in mind is that you want to develop unique combinations of abilities. You need to develop unique combinations of talents which give you a comparative advantage over other people. *Do what you can do better than anybody*; to mangle a quote from Lincoln, nobody can be better than everybody all of the time, but anybody can be better than everybody some of the time.

In my opinion the reason most people fail to do great research is that they are not willing to pay the price in self-development. Say some new field opens up that combines field X and field Y. Researchers from each of these fields flock to the new field. My experience is that *virtually none* of the researchers in either field will systematically learn the other field in any sort of depth. The few who do put in this effort often achieve spectacular results.

Finally, a note on how to go about developing some new research strength. A mistake I'm prone to make, and I know some others are as well, is to feel as though some degree of completeness is required in understanding a research field. In fact, in any given research field there are usually only a tiny number of papers that are really worth reading. You are almost certainly better off reading deeply in the ten most important papers of a research field than you are skimming the top five hundred.

These ideas carry over to the problem of staying current in your fields of interest: I believe that you can stay quite current by (a) quickly skimming a great deal of work, to keep track of what is known, and what sort of problems people are thinking about, and (b) based on that skimming, picking a dozen or so papers each year to read deeply, in the belief that they contain the most important research results of the year. This is not the only deep reading you'll need to do; you'll also need to do some which is related to the immediate problems that you're working on. But you certainly should do some such deep reading.

Develop a high-quality research environment

There is a considerable amount of research showing that people consistently underestimate the effect of the environment on personal effectiveness. This is particularly important in an academic environment where there are usually many short-term social pressures that are not directly related to research effectiveness – teaching, writing letters of recommendation and referee reports, committee work, academic politics. By contrast, in most institutions there are few short-term social pressures to do great research work.

Some of the highest-leverage work you can do involves improving your environment so that social pressures work *for* you as a researcher, rather than against you. Discussing this in detail would require another essay of length at least equal to that of the present one, but I will make a few remarks.

The first is that improving your environment is something anyone can do; students, in particular, often underestimate the magnitude of the changes they can bring about. Anyone can start a seminar series, develop a discussion area, create a lounge, organize a small workshop, or organize a reading group. Furthermore, although all these things are hard to do well, if you're willing to do critical evaluations, experiment and try radical changes, preferably in partnership with equally committed people, things are likely to improve a great deal.

Second, institutions have long memories, so changes that you make in your environment will stick around for a long time. This means that once something is working well, chances are it'll continue to work well without much help from you – and you can move on to improve some other aspect of your environment. Furthermore, each positive change you make actually improves your leverage with other people. I've known undergraduate students who had made so many creative positive contributions to their departments that their influence with canny senior faculty was comparable to the influence of other senior faculty.

The creative process

The problem-solver and the problem-creator

Different people have different styles of creative work. I want to discuss two different styles that I think are particularly useful in understanding the creative process. I call these the *problem-solver* and the *problem-creator* styles. They're not really disjoint or exclusive styles of working, but rather idealizations which are useful ways of thinking about how people go about creative work.

The problem-solver: This is the person who works intensively on well-posed technical problems, often problems known (and sometimes well-known) to the entire research community in which they work. The best problem-solvers are often extremely technically proficient and hard-working. Problem-solvers often attach great social cache to the level of *difficulty* of the problem they solve, without necessarily worrying so much about other indicators of the importance of the problem.

The problem-creator: This is a rarer working style. Problem-creators may often write papers that are technically rather simple, but ask an interesting new question, or pose an old problem in a new way, or demonstrate a simple but fruitful connection that no-one previously realized existed.

Of course, the problem-solver and the problem-creator are idealizations; all researchers exemplify both styles, to some extent. But they are also useful models to clarify our thinking about the creative process. One distinction between the two styles is how proactive one is in identifying problems, with the problem-solver being much more passive, while the problem-creator is extremely proactive. By contrast, the problem-solver needs to be much more proactive in developing their problem-solving skills. Both styles of research can be extremely successful.

Problem-solvers have numerous social advantages in research, and for that reason I believe they tend to be more common. In particular, it is relatively easy to recognize (and then reward) people who solve problems that are of medium or high levels of difficulty. This has rewards both in terms of the immediate esteem of one’s peers – physicists love to trade legends about brilliant colleagues who immediately see through to the solution of some difficult problems or another – and also in the hunt for jobs and other tangible forms of recognition. It takes more time (and thus can be more difficult) to recognize people whose work is technically rather simple, but whose questions may eventually open up whole new lines of enquiry.

The advantage in being a problem-creator is that there is a sizeable comparative advantage in opening up an entirely new problem area, and thus being the first into that problem area. You can work hard to get a basic foundation in the skills needed in that problem area, and then clean up many of the fundamental problems.

The skills of the problem-creator

Our training as physicists focuses pretty heavily on becoming problem-solvers; we tend not to get much training as problem-creators. One reason I’m discussing these two working styles at some length is to dispel the common idea that creative research is necessarily primarily about problem-solving. It’s true that many people have very successful research career as problem-solvers. But you can also consciously decide to invest more time and effort into developing as a problem-creator. I now describe some of the skills involved in problem-creation.

Developing a taste for what's important: What do you think are the characteristics of important science? What makes one area thrive, while another dies away? What sorts of unifying ideas are the most useful? What have been the most important developments in your field? Why are they important? What were the apparently promising ideas that didn't pan out? Why didn't they pan out? You need to be thinking constantly about these issues, both in concrete terms, and also in the abstract, developing both a general feeling for what is important (and what is not), and also some specific beliefs about what is important and what is not in your fields of interest. Richard Hamming describes setting aside time each week for "Great Thoughts", time in which he would focus on and discuss with others only things that he believed were of the highest importance. Systematically setting aside time to think (and talk with colleagues) about where the important problems are is an excellent way of developing as a problem-creator.

On this topic, let me point out one myth that exerts a powerful influence (often subconsciously) on people: the idea that difficulty is a good indicator of the importance of a problem. It is true that an elegant solution to a difficult problem (even one not *a priori* important) often contains important ideas. However, I believe that most people consistently *over rate* the importance of difficulty. Often far more important is what your work enables, the connections that it makes apparent, the unifying themes uncovered, the new questions asked, and so on.

Internal and external standards for what is important: Some of the most thought-provoking advice on physics that I ever heard was at a colloquium given by eminent physicist Max Dresden. He advised young people in the audience not to work towards a Nobel Prize, but instead to aim their research in directions that they personally find fun and interesting. I thought his advice quite sound in some regards: for some people it is extremely tempting to regard external recognition as the be-all and end-all of research success, and the Nobel Prize is perhaps the highest form of external recognition in physics. Dresden is right, in the sense that working with a primary goal of winning a Nobel Prize would be pointless and degrading; far better to work in an area one personally finds enjoyable.

On the other hand, the Nobel Prizes are usually given for very good reasons: they reward some of the most interesting work in all of physics. There is, admittedly, a political element, with certain fields being favoured, and so on. Nonetheless, imagine a world in which one of these discoveries had *not* been awarded a Prize for some reason. Would you be proud to have your name associated with that discovery, even so, and regard the work on it as time well spent? In every case I can think of, that certainly is the case for me, and I suspect it's true for most other physicists.

I believe this highlights an interesting point about what makes something interesting and important. A person working toward a Nobel Prize or some other form of external recognition has, in some sense, decided to abdicate their personal decision about what is important and interesting. The external community of physicists (in this case, represented by the Nobel Committee) is what makes their decision: if it might win a Nobel, it's important.

Balancing this observation, this is not to say that your decision about what is interesting and important should be yours along. People who work in isolation rarely end up making contributions that are all that significant. Your decision about what is important should be *informed* by others: talk to your peers, find out what they think is important, look in the textbooks and history books and biographies, and, yes, look at what wins prizes (of all sorts).

But at the end of the day you've got to form your own independent standards for what is interesting and important and worth doing, and make judgments about where you should be making a contribution, based on those standards. I think better advice from Dresden would have been to aim to produce work of the highest possible caliber, but according to what *you* have come to believe is important.

Exploring for problems: Obviously, all researchers do some of this. For the problem-solver, the process of exploring for problems often works along the following lines: keep moving around, looking for problems that you consider (a) well-posed, or able to be well-posed after some work on your part, (b) likely to fall within a reasonable time to the arsenal of tools at your disposal (perhaps with some small expansion of that arsenal), and (c) below some minimum thresholds of interest and difficulty. Once you've found a problem of this sort, you work hard on the problem, solve it, and publish.

Problem-creators may be rather more systematic about exploring for problems. For example, they may occasionally set time aside to survey the landscape of a field, looking not just for problems, but trying to identify larger patterns. What *types* of questions do people in the field tend to ask? Can we abstract away patterns in those questions? What other fields might there be links to? What are the few most important problems in the field? Problem-creators set aside time for doing this kind of systematic exploration, and do it in a disciplined way, often with feedback from others.

Surveying the landscape can be particularly revealing. A lot of people work in fashionable subfields of a larger field primarily because there are lots of other people working in that subfield. The problems they work on may be technically complicated, especially after a few years, when the most basic questions have been answered. This is compensated by the fact that it's extremely comforting to work within a field where there is a standard narrative explaining the importance of the field, some canonical models for what problems are interesting, and a willing audience of people ready to appreciate your work. In addition, working in such subfields gives younger people a chance to show off their technical prowess (sometimes, not unlike elk spoiling for a fight) to peers in a position to recommend them for valuable faculty positions.

Getting ahead of the game: There are many important problems, and sometimes an entire field comes to some agreement about what is important: proving the Riemann Hypothesis, or understanding high temperature superconductivity. Sometimes, however, there is a problem either not appreciated at all, or only dimly appreciated, that is equal in importance to such gems. Consider the creation of the scanning tunneling microscope – the basic idea had been around for years, yet nobody had ever seriously tried to build the device. The inventors put it together on a shoestring, and created one of the major tools of modern physics. Or consider David Deutsch and Richard Feynman's creation of the field of quantum computing, by framing the right questions ("What would a quantum mechanical computer be capable of?" and "Would it be faster than a classical computer?"). One of the big ways you can get ahead as a researcher is by identifying and then solving problems that are important, but perhaps not terribly difficult, ahead of everyone else.

Identify the messes: In a nice article about how he does research, physicist Steven Weinberg emphasized the importance of identifying the messes. What areas of physics appear to be a state of mess? Funnily enough, one of the signs of this can be that it's very hard to understand. For a long time – and to some extent this persists today – physics texts on general relativity were very difficult to understand. The tensor calculus in them was often confusing and difficult to understand. There was a good reason for this: the basic definitions in the subject of differential geometry, although laid down in the 19th century, didn't really reach their modern form until the mid part of the twentieth century, and then took considerable time to migrate to physics. The reason a lot of the discussion of tensor calculus in physics texts is confusing is because, very often, it is *confused*, being written by people who don't have quite the right definitions (meaning, in this case, simplest, most elegant and natural) in mind.

When you identify such a mess, the natural inclination of many people is to shy away, to find something that is easier to understand. But a field that is a mess is really an opportunity. Chances are good that there are deep unifying and simplifying concepts still waiting to be understood and developed by someone – perhaps you.

The skills of the problem-solver

As I've already said, our technical training as physicists focuses a lot more on problem-solving than problem-creation, so I'm not going to say a lot about the skills needed to be a problem-solver. But I will make a few general remarks that I find helpful.

Clarity, goals, and forward momentum: In my opinion, there is little that is more important in research than building forward momentum. Being clear about some goal, even if that goal is the wrong goal, or the clarity is illusory, is tremendously powerful. For the most part, it's better to be doing something, rather than nothing, *provided*, of course, that you set time aside frequently for reflection and reconsideration of your goals. Much of the time in research is spent in a fog, and taking the time to set clear goals can really help lift the fog.

Have multiple formulations: One of the most common mistakes made by researchers is to hold on very closely to a particular problem formulation. They will stick closely to a particular formulation of a problem, without asking if they can achieve insights on related problems. The important thing is to be able to make some progress: if you can find a related problem, or reformulate a problem in a way that permits you to move forward, that is progress.

Spontaneous discovery as the outcome of self-development: For me this is one of the most common ways of making discoveries. Many people's basic research model is to identify a problem they find interesting, and then spend a lot of time working on just that problem. In fact, if you keep your mind open while engaging in exploration, and are working at the edge of what is known, you'll often see huge opportunities open wide in front of you, provided you keep developing your range of skills.

Working on important problems

It’s important that you work towards being able to solve important problems. This sounds silly, but people don’t do this for any number of reasons. I want to talk a little about those reasons, and how to avoid them.

Reason 1: Lack of self-development. Many people don’t spend enough time on self-development. If you stop your development at the level which resulted in your first paper, it’s unlikely you’ll solve any major problems. More realistically, for many people self-development is an *incidental* thing, something that happens while they’re on the treadmill of trying to solve problems, generate papers, and so on, or while teaching. While such people will develop, it’s unlikely that doing so in such an *ad hoc* way will let them address the most important problems.

Reason 2: The treadmill of small problems. Social factors such as the need to publish, get grants, and so on, encourage people to work only on unimportant problems, without addressing the important problems. This can be a difficult treadmill to get off.

My belief is that the way to start out in a research career is by working primarily on small and relatively tractable problems, where you have a good chance of success. You then continue the process of self-development, gradually working up to more important problems (which also tend to be more difficult, although, as noted above, difficulty is most emphatically not the same as importance). The rare exception is important problems that are also likely to be technically easy; if you’re lucky you may find such a problem early in your career, or be handed one. If so, solve it quickly!

Even later on, when you’ve developed to the point that you can realistically expect to be able to attack important problems, it’s still useful to tackle a mixture of more and less important problems. The reason is that tackling smaller problems ensures that you make a reasonable contribution to science, and that you continue to take an active part in the research community. Even Andrew Wiles continued to publish papers and work on other problems during his work on Fermat’s Last Theorem, albeit at a rather low rate. If he had not, he would have lost contact with an entire research community, and losing such contact would likely have made a significant negative difference to his work on Fermat’s Last Theorem.

Reason 3: The intimidation factor. Even if people have spent enough time on self-development that they have a realistic chance of attacking big problems, they still may not. The reason is that they have a fear of working on something unsuccessfully. Imagine Andrew Wiles feeling if he had worked on Fermat’s Last Theorem for several decades, and completely failed. For most people, the fear of ending up in such a situation is enough to discourage them from doing this.

The great mathematician Andrei Kolmogorov described an interesting trick that he used to get around this problem. Rather than investing all his time and effort on attacking the problem, he’d put the problem into a larger context. He’d announce a seminar series in which he’d lecture on material that he thought would be related to the problem. He’d write a set of lecture notes (often turning into a book) on material related to the problem. That way, he lowered the psychological pressure on himself. Rather than investing all his effort in an attack on the problem – which might ultimately be a complete waste of time – he knew that he'd produce something of value. By making the research process part of a larger endeavour, he ensured that the process was a success no matter how it came out, even if he failed to solve the problem, or was scooped by someone else. It's a case of not putting all of one's psychological eggs in one basket.

Richard Feynman described a related trick. He said that when he started working on a problem he would try to convince himself that he had some kooky insight into the problem that it was very unlikely anybody else had. He admitted that very often this belief was erroneous, or that, even if original, his initial insight often wasn't very good. But he claimed that he found that he could fool himself into thinking that he had the "inside track" on the problem as a result, and this was essential to getting up the forward momentum necessary to really make a big dint in a difficult problem.

Committing to work on an important problem: For the difficult problems, I think commitment is really a process rather than a moment. You may decide to prepare a lecture to talk about a problem. If that is interesting, you enjoy it, and you feel like you have some insight, you might decide to prepare a few lectures. If that goes well, perhaps you'll start to prepare more lectures, write a review, and maybe make a really big contribution. It's all about building up more and more insight. Ideally, you'll do this as part of some larger process, with social support around you.

People who only attack difficult problems: There is a converse to the problem I've been talking about, which is people who are only interested in attacking problems that are both difficult and important. This affliction can affect people at any stage of their career, but it manifests itself in somewhat different ways at different stages.

In the case of the beginner, this is like a beginning pole vaulter insisting on putting the bar at 5 meters from the time they begin, rather than starting at some more reasonable height. Unless exceptionally pigheaded, such a person will never learn to vault 5 meters successfully, simply because they will never learn anything from failure at a more realistic starting height. This sounds *prima facie* ridiculous, but I have seen people burn out by following exactly this strategy.

The case of the more experienced researcher is more difficult. As I've emphasized, once you've reached an appropriate level of development I think it's important to spend *some* time working on the most important problems. But if that's all you do, there are some very significant drawbacks. In particular, by attacking only the most important and most difficult problems an experienced researcher (a) takes themselves out of circulation, (b) stops making ongoing contributions, (c) loses the habit of success, and (d) risks losing morale, which is so important to research success. I think the solution is to balance one's work on the more and less important problems: you need to schedule time to do the more important stuff, but should also make sure that you spend some time on less high-risk activities.

In both cases, the explanation is often, at least in part, intellectual macho. Theorists can be a pretty judgmental lot about the value of their own work, and the work of others. This helps lead some into the error of only working on big problems, and not sometimes working on little problems, for the fun of it, for the contact it brings with colleagues, and for the rewarding and enjoyable sense of making a real contribution of some significance.