GOOD AND NOT-SO-GOOD IDEAS IN RESEARCH.
A TUTORIAL IN IDEA ASSESSMENT AND GENERATION

Lennart Sjöberg

Center for Risk Research
Stockholm School of Economics
lennart.sjoberg@hhs.se

1 June 2003

Good and not-so-good ideas in research.
A tutorial in idea assessment and generation

Lennart Sjöberg[1]
Center for Risk Research
Stockholm School of Economics

Abstract

This starting point of this paper is the difficulty for many graduate students to ever finish a Ph. D. and the graduate student s need for good research ideas. According to Swedish data, only 20 percent of those who start as graduate students in the social sciences ever finish. One crucial problem is probably the lack of ideas and awareness of how to foster creativity. What *are* good ideas? Ideas are assessed all the time, by journal editors and referees, by research councils and thesis supervisors. Yet, there is little reflection on the question of what constitutes good ideas, and bad. Philosophy of science is briefly discussed in the paper, with examples from behaviorism and social constructivism, and it is argued that it provides no good basis for generating and assessing good ideas. The paper then proceeds by discussing examples of good, and not-so-good ideas. The notion of good bad ideas is introduced and an example is given. The paper then discusses several questions. Which characteristics of research ideas make them good, and which make them less good? How does one go about creating good ideas? What sort of research environment is most likely to stimulate the growth of good ideas? What are the processes which kill creativity in research? Creativity requires an open mind and strong interest. It takes time, sometimes much time, before good ideas arrive. Yet, success in a research career requires publications, and a brief section discusses how to plan one s publications and have them accepted by journal editors. Finally, Swedish experience with an attempt to reform graduate education is discussed. It is pointed out that administrators inclinations to favor concrete policy measures such as financing do not necessarily lead to the desired consequences of increasing completion rate and decreasing study time. Another approach must be applied: the fostering of creativity.

Key words: graduate education, creativity, assessment of research ideas

---

Introduction

The occasion for writing the present paper arose in graduate education. I have much experience in supervising the research of doctoral students, yet before writing this paper I had reflected fairly little on how it should be done. What are good and what are bad ideas, and how do you get and develop them? This paper is, however, not written from the standpoint of supervision, but mostly from the standpoint of the student. I assume throughout that the student strives to generate his or her own ideas, an assumption which is not always true. So, clearly there are several topics I could not deal with here which are important and related to the general theme of how to launch a research career. Hopefully, the present paper is a piece of the puzzle.

The paper is attuned to conditions in graduate study in Sweden. Some details may be specific to that country, but in most respects I think the discussion is general enough to be of some interest in other contexts as well. In addition, the discussion is geared towards psychology and some parts of it may be special to that discipline, or to a group of social science disciplines.

Are you looking for a good idea for a thesis in psychology? Is it hard to come up with one? This paper will tell you why, point to examples of good ideas, as well as examples of not-so-good ideas and what distinguishes them. The paper discusses how ideas can be evaluated. It will also try to tell you *how* to get good ideas, and how to develop them.

A basic assumption is that you *are* looking for good ideas for research. This is not so obvious as it may sound. The current idea about Ph.D. work in Sweden assumes that a thesis is done in two years. Very few manage to do that and the main problem is probably lack of workable ideas. Thesis supervisors could of course provide such ideas[2]. Yet, students want to work on their own ideas. It takes a very long time for most students to develop a theme for a thesis, and it is not a question of hard work . It is a question of waiting for closure. I am convinced that this is largely a waste of time. It is somewhat absurd that it could take 20 years to complete a thesis, which has been known to happen. Some students never make it, in fact only 20 percent of students accepted for graduate study in the social sciences in Sweden ever take the Ph. D. degree[3]. The process can be much faster, and much more efficient, provided there is some insight into what it takes to generate ideas. The purpose of the present paper is to serve as a basis for discussions which will help students work on the process, understand what to avoid and what to strive for, and learn where and how workable ideas can be found.

The process of thesis work is largely a question of mobilizing the right level of motivation. This is no simple thing. Many students have an exaggerated respect for what it takes to be a

---

[2]. It has been pointed out to me that some doctoral students have very small chances to bring up and work on their own ideas, since they are employed in their supervisor s projects. If this is true, it is a very different approach to graduate education which is not discussed in the present paper.

[3]. These data pertain to the situation before 1998 when a reform was implemented; the effects of that reform, if any, have not yet been documented.

researcher, and what it takes to pass the ultimate academic test, the Public Defense. I have known students who had a high level of talent and motivation for research, yet could not face the Public Defense. We touch here on personality problems that a thesis adviser cannot treat and has no responsibility for. He or she is no therapist, and should not try to be one. But one thing that the supervisor can do is to boost morale and motivation by being supportive and positive and genuinely interested in the student s work, and give reasonably fast feedback on manuscripts. The supervisor at times also has his or her own demons. The perfectionistic supervisor will not let the student finish a thesis until it is 2 or 3 times above what is reasonably required. In this way, the supervisor avoids being criticized by colleagues for demanding too little from students. The students pay the price.

Most examples of good and bad ideas in the present paper are from my own work, which is currently concerned with risk perception and with emotional intelligence. A previous monograph in Swedish discussed some of the issues of research strategy in more detail (Sjöberg, 1977a). In other papers I have discussed related issues of substance and methodology (Sjöberg, 1981, 1983, 1987, 2000a). The present paper is not an attempt to summarize this previous work of mine, but more of a progress report.

Let the reader be aware that the author has spent more time than he should have, in assessing research ideas. This has been a major part of my professional life for 30 years, as a thesis adviser and also as a member of research councils and a referee for councils and scientific journals. This kind of work is largely about assessing the value of ideas. Strangely enough, there is very little explicit discussion in these contexts about just what is good, and what is bad, and why.

Take as an example the aspect of theoretical basis . I heard many times this criterion to be proposed as a major aspect on which to assess research proposals. It sounds good, yes? Well, it all depends on what is meant. It turned out that almost no applications for research funding contained any critical remarks about theories, compared theories or suggested to work for the critical, empirical, assessment of theories, or to create new theory. None of all that! Instead, a passing grade was awarded whenever the applicant cited the right current theories, and correctly referred what these august scientists had claimed. A theoretical basis turned out to involve the use of theory as a justification and showing that one had read the current standard references. There is nothing wrong with such reading, of course, but if research is to make progress, reading must be critical. It must be critical in the scientific sense, it must question the theories on empirical and conceptual grounds, not on ideological ones.

A few words about what will *not* be treated in the present paper are in order. Prescriptions for research often emanate from theory of science and philosophy. Researchers in psychology have indeed been very much pre-occupied by the question of what constitutes scientific knowledge in psychology, and decades have been wasted on programmatic research in the belief that it was scientific . The best example is behavioristic learning research 1910-1960, 50 years! During this time, the theoretical basis of the study of behavior was believed to be found in studies of animal learning, presumably because even simple animals such as rats exhibit the same basic processes as human beings, but can be studied in a more straightforward, experimental manner. Some principles of behavior can perhaps be established in this way, but the whole spectrum of human learning and information

processing is missed. The reasons for the - to some - enormously attractive idea of animal models were partly to be found in a behavioristic philosophy of science, and it is my conviction that good research ideas are very unlikely to be found in that environment. Theory of science has other merits, but it does not function as an idea generator, nor is it usually intended as one. Let s look at another example.

A second, and more current, example is social constructivism (Hacking, 2000). This is just about the opposite of behaviorism and adherents of this approach to social and behavioral science claim that all there is, is social construction. However, this is both true and false. It is true that all our perceptions and constructs regarding the external world in all its aspects are constructions, many or most of them perhaps even social constructions. We inherit them from the culture we live in and seldom add much in the way of idiosyncracies. At the same time, these constructs usually refer to an external world, there is something out there . Take political attitudes. Most people have political attitudes, e. g. in the form of likes and dislikes of political parties or leaders. These attitudes can be studied by means of surveys or interviews and in other ways as well, and we can reach certain conclusions about them. These conclusions are social constructions, but they are not arbitrary or equally good no matter what they contain. Some conclusions are closer to reality, and there is a reality out there . Research can improve and accumulate better and better knowledge about people s attitudes, even if nobody would seriously argue that we can arrive at the final truth about *anything* in social science. Social constructivism often denies these banal truths about the external world, and a conceptual world, and mixes them up, ending up in a totally nihilistic standpoint where anything goes (Windschuttle, 1996). If there are no criteria, except perhaps linguistic elegance and persuasive power, to assess the validity of research, it is hard to believe that there can be progress. Indeed, the constructivist will say that the very idea of progress is naive. But in my view, good ideas have a relationship to realities which are independent of the ideas. Hence, I see little use for social constructivism. Windschuttle s penetrating analysis should be consulted in case you still believe in social constructivism (Windschuttle, 1996). Sokal exposed the emperor as naked, in a very entertaining manner (Sokal & Bricmont, 1998).

A third example, from psychology, is related to social constructivism. Smedslund has argued, for many years, that psychological research tends to be tautological, that the hypotheses tested must be true by definition (Smedslund, 1991). The point is well taken at times, but should not be exaggerated. There is much empirical research in psychology where it is not relevant (Sjöberg, 1982, 1999a). Smedslund s achievement is that he has made researchers observant of the dangers of tautologies. It is interesting to reflect on whether the same argument can be made in other disciplines.

What is good ?

The present paper talks about good ideas, not merely ideas. It is therefore necessary to reflect on the meaning of good . *A good idea is new and original and it is related in an interesting way to theoretical developments in the field and/or to practical problems.* Usually, a purely practical problem would not qualify as a good research idea, there need to be some elements of general interest beyond a concrete application. There must also be intellectual

challenge involved. This factor is essential for study interest (Sjöberg, 1985, 1997; Sjöberg & Dahlstrand, 1987). In addition, I do not deal with purely theoretical developments in the present paper, the ideas discussed must have, at least in principle, some empirical substance[4].

A good idea is also a workable idea. It must be possible to make something of it, it should not be too difficult to develop into an empirical application. The paper also deals with not-so-good ideas and it suggests why such ideas are less promising. The reason for taking that approach as well is that many students often come up with such less than useful ideas, on the basis of folk psychology and common sense.

The source of a good idea is usually a combination of knowledge of the literature and practices in a subfield, specific theories and principles, common sense and the researcher s own phenomenology. But beware, some ideas that seem good, aren t! An example will serve as a starter.

A bad idea

Psychology students very often suggest, as a research topic, to study the effects of some kind of psychotherapy. What s wrong with that? There are many questionable aspects which need to be discussed.

First, it is a superficial idea, usually not motivated by knowledge of the cutting edge of therapy research. It is motivated by the student s interest in a form of therapy, not interest in the scientific basis of it. This is the wrong place to start research. One should start from the therapy research literature, which is very extensive, and derive research topics from there. This, in turn, is a big order and requires extensive and discriminating reading and skilful judgment, not likely to be a possible and realistic goal for the student.

Second, therapy research is quite hard to do well, and seldom within the resources available to a student. It needs lengthy observations, and the cooperation of professionals who will have to do a considerable amount of work in order to comply with the design requirements of the study, and who may also be defensive about what they are doing. They are not necessarily charmed by the idea of having a student from some university, and his or her professors, peek into what they do for a living. Also, many therapists are hostile towards systematic research and what it brings with it in terms of requirements for control groups and even measurement. They feel that what they do is very important and valuable but they doubt that it can be measured. A further practical complication is that patients need to form some homogeneous diagnostic group. Most patients are not clear examples of such a group and it may therefore take quite a long time before a group which is large enough is collected.

Third, the simple question Does x-therapy have any effect? leaves me, and many thesis supervisors, cold. There are literally thousands of studies of this type, and even if a new minor project might add a piece of the puzzle, it is bound to be very minor. It also has no or

---

[4]. The reason is mainly that graduate students research rarely is advanced enough to contribute to theory. It takes a long time for a researcher to reach that stage.

very little intellectual challenge to it - it is too much of a purely practical question. Intellectual challenge is a cornerstone of scientific interest (Sjöberg, 1997).

The example illustrates how a bad idea arises out of non-scientific interests. There is nothing wrong with such interests *per se*, they are just relatively unlikely to bear fruit in science. Exceptions happen and will be discussed in a later section.

Why is it so hard to be creative?

Can anyone be creative? Can anyone come up with good research ideas? Probably not, but many can do it, including many who doubt their own capacity in this respect. There are several obstacles, though.

First, there is the problem of *excessive respect for and deference to authority*. This is a very common problem. Social science researchers (in Sweden, but probably everywhere) tend to have a very high regard for theory which is fine up to a point, but not when excessive. The great and well-known developers of social or behavioral theory are read, their messages hopefully understood (when that is possible) and the student then goes on to repeat what those famous people have written, only in a less clear and stimulating manner. He or she offers no critical remarks, no new ideas and no plans for crucial experiments or other types of empirical studies which could throw some light on the validity of these theories. Validity? Do I mean they may actually be *wrong*? Yes, I do. The whole point of research is that of formulating and testing theories, and improving on them whenever they fail empirically. Naturally, any empirical test is contingent on the validity of the procedures used in the test, such as scales used for measuring crucially important concepts. These scales can be always criticized in themselves. Research is always full of uncertainties, which is one reason why some people rightly prefer to do something else. The student needs to be less intimidated by the international luminaries in the field, and to feel a healthy lack of respect. They are all human, and make many errors.

The idea of empirical tests is less innocent than it sounds. For many students it is not a natural way of thinking. They prefer to judge a theory on other grounds, such as the persuasive power of the theorists or how colorful applications of the theory are, and even how much they would wish that things were the way the theory asserts them to be.

Second, there is a problem of seeing the forest, in spite of all the trees . The student risks to be *involved in a large number of problems of detail, which may be challenging but are not the most crucial ones*. Some of these problems are statistical. How should one carry out a statistical analysis of one kind or another? Many problems of this type pertain to significance testing, which is a class of data analysis procedures which has probably caused much more harm than good to psychological research (Schmidt, 1996; Sjöberg, 1999b). There are two reasons for this. The first, and most important, is that researchers have been blinded to the much more important requirement of establishing the *size* of effects, not only if they are non-

6

random in the sense that a random process is unlikely to have caused them[5].

The second reason why significance testing is destructive is that it takes time and energy from more important issues. How to test significance is a thorny topic by itself, and if pursued it will lead the researcher into a foggy marsh where only a fool would rush in. How important is the common assumption of normal distributions? Nobody seems to know anymore, although I was taught, in my graduate studies, that statistical tests were robust. Well, that is not true any more. They may be, or maybe not.

This point is related to measurement more generally. Here is another aspect where my graduate study was misleading. I was taught that measurement level is enormously important, and that psychology must strive to develop new and better measurement methods (Stevens, 1951). What a waste of time that was! The simple methods we have serve us well, and even if they have not been proven to yield interval measurements they are in all likelihood good enough approximations. Anderson has shown, in a very extensive research program using simple categorical rating methods, that data have patterns supporting this assertion (Sjöberg, 1966; Anderson, 1982; Anderson, 1996; Sjöberg, 1994). In addition, the scaling methods devised by Stevens and Ekman, among others, in the 1950's and 1960's (Ekman & Sjöberg, 1965), turned out to carry with them their own very difficult problems (Sjöberg, 1971). There was nothing straightforward, after all, about direct psychological measurement, as Ekman called it (Ekman, 1958).

The philosophical analysis of fundamental measurement is theoretically interesting but seems to generate very few interesting empirical applications (Cliff, 1992). Unfortunately, some statisticians currently stress the measurement issues. Since there is no good alternative to the simple rating methods now used all over psychology, their criticism leads nowhere but may deter researchers from using the simple and effective methods which do exist.

Mathematical analysis is unsuitable to psychology because phenomena are prone to error variability between individuals (usually enormously large) and also within individuals, across trials or occasions (Sjöberg, 1977a). This error variability has little in common with random processes but is almost always systematic. Simple linear models will suffice to account for data in most situations, the real issue is not the weak deviations from linearity that can be detected but how to get a better over-all fit, in other words how to increase correlations or effect sizes (Sjöberg, 2003a).

Third, there is the problem of *too little or too much self-confidence*. Those who have little self-confidence just cannot believe that there own ideas are worthy of bringing forward and working on. This is a very common attitude. The situation is made worse by the academic culture which often encourages an overly critical attitude. The academic seminar is a standard setting for establishing a pecking order, for people to criticize each other from top to bottom.

---

[5]. We knew that from the beginning, anyhow, and all that statistical hypothesis testing does is to establish that the study design was powerful enough to let us discover it. Although we often talk about randomness, we never really mean that events are without a cause (Sjöberg, 2002). We always impose meaning and causal structure on events.

7

There may be a gender factor involved, since men are more likely to assert themselves in this way. No matter what, it discourages many students and intimidates them from developing their own creative force. Almost all people have such a capability, but if discouraged, many will just give up.

It is obvious that the academic seminar *must* be a place for venting critical thoughts. This could be done in a less threatening manner than what is often the case, however, and there should be more awareness and reflection on the great risks for motivation. It is indeed possible that a seminar leader who is a cordial and not terribly smart or knowledgeable person is better for the students than a highly efficient and sharp person who immediately detects any weaknesses. The students will get more information from the latter type, but they will also risk losing their zest. From the former type, they get positive feedback which will sustain them during the first difficult years of a research career, and function as emotional support when they meet with the icy winds blowing on the academic tundra. I refer, for example, to the anonymous referee s reports they will get from international journals. They are sometimes dripping with poison and very mean. There is no way around this, however, if the researchers are to get their work read and attended to by the international community (Sjöberg, 1976, 1977b)[6]. Experience of research is not enough, there must also be feedback if a researcher is to develop a high level of achievement (Ericsson, 1999).

How about too much self esteem? This risk has been discussed by Baumeister et al. (Baumeister, Campbell, Krueger, & Vohs, 2003). In my experience, it is not uncommon that students have enormously high aspirations for their work. They aim for the sky. They are hypercritical of the work of others, and also of their own ideas. They are mesmerized by the idea that they will revolutionalize science. I don t know from where they get such ideas. They are seldom based on realistic self knowledge, for these are mostly people who achieve very little indeed. They do not conform to deadlines and finishing a thesis will take not a year or two extra but a decade or two above the stipulated time for completing the work. In addition, when they finally complete a thesis, if at all, it is seldom particularly interesting or creative. It may be water tight , however, for many of these students are compulsive perfectionists. Striving for the perfect is very dangerous because attention to details carries the risk of *only* attending to details. Compulsive perfectionism is extremely counter-productive and it requires personality change to get rid of it (Sapadin, 1999). The thesis adviser is no therapist, however, and cannot take on that task. These people also see themselves as models of seriousness och ambition and it takes many years before they realize that they spend their lives on trivial details and get nothing of interest accomplished.

The fourth obstacle to creativity to be mentioned here is the *reluctance to be integrated in an intellectual community.* Some students see graduate studies and research as just another job . A recent survey at my own school talks about research and graduate study as a burden . It is something to do, at most, between 9 and 5, 5 days per week. There should be long

---

[6]. It is common in some social science disciplines in Sweden that this step is only seldom taken, and that local groups prefer not to have their work assessed by outsiders. Some sociology departments are like this, and it pertains even more to departments of business administration (Sjöberg, in press) . While this certainly makes for avoiding some unpleasant criticism, it also makes the whole research effort meaningless because very few people will pay any attention to it.

holidays and vacations as well. But research does not function like that. Someone who sees his or her research in this way should probably be doing something else. If there are to be results, research cannot be turned on and off at will. It integrates with the whole person, and is a way of life. The researcher does not care what time of the day it is, or if others are lying on the beach. His mind is forever busy with his interests, and the interests will never let him go. This may sound like a one-sided life and it is. The great scientists such as Sir Isaac Newton have testified to having had this lifestyle (Tiner, 1981). Everyone cannot be a new Newton, but that is not the point. The point is this: if research is seen as just another job, how can you ever hope to accomplish anything beyond routine work? Research is a highly personal activity, driven by inner motives of interest and the joy of getting new ideas, getting *your own* ideas, and seeing them reach the international scientific community. The researcher also takes interest in the wider environment, in his group of colleagues, even if they work on other topics than his own. He or she finds time for that. In this way, a constructive environment is created and sustained.

I now turn to examples of good ideas. If nothing else, they will give the reader an idea of what *I* think are good ideas. If he or she does not agree, this may be a fruitful ground for a debate.

Examples of good ideas

Example 1: Test motivation

Here is a first example of an idea which may seem to be good, but there is perhaps some hidden weakness in it.

People take ability tests and the purpose is to measure their abilities. This has been going on for about 100 years. A certain level of modest success has been achieved in this way. Spearman formulated the *g*-factor theory saying that there was a general intelligence factor common to all these tests. Thurstone, Guilford and Cattell followed in his footsteps. With each new generation, the number of factors increased, but the basic technology of test construction was unchanged and tests looked very much the same. The main development was a remarkable proliferation of factors, and tests, but little improvement in understanding of mental abilities and no improvement to speak of in practical achievements.

Three radical ideas for improvement in ability testing now suggest themselves: to measure test motivation, to investigate cognitive processes in detail and to switch from traditional test items to the measurement of practical abilities or competencies (McClelland, 1973, 1998). The first is still unchartered territory, the second is a junk yard of failed hopes and the third is a story of merely modest success.

It is hard to understand why ability testers have not long ago devised measures of test motivation. In the neighboring field of self-report scales of personality, impression management is measured, at least since the beginning of the 1960's (Crowne & Marlowe, 1960). New developments are still being published (Paulhus, 1991, 1998; Paulhus, Bruce, & Trapnell, 1995; Paulhus & Reid, 1991). Faking is measurable and has been shown to fully explain almost all of the enormous differences of self-report data in a real, high-stakes testing situation and a situation where all data were guaranteed to be completely anonymous and

would have absolutely no consequences for the testees (Sjöberg & Engelberg, 2002). Yet, motivation in ability testing remains to be measured. How should we proceed with this good idea?

It will not do to ask people how much of an effort they exerted, at least not as the sole measure. Other ideas must be developed. How should it be done? Here is a suggestion. In a multiple choice test a little motivated testee might be using stereotypical response patterns based on position of the selected alternative and previous choices, such as alternating responses. If the correct responses are strictly randomly distributed across response alternatives, any non-randomness of incorrect responses would reveal a lack of effort.

The idea could be tested against criteria of self-reported effort, experimental manipulations of motivation by means of incentives enticing people to work hard, and in prediction where a motivation score would contribute to the predictive value of the test. A given score should be a stronger sign of underlying ability if it was achieved with little effort than if effort to achieve it was very great. (On the other hand, the fact that someone invested only little effort could mean that he or she has other undesirable characteristics as well...).

What s wrong with this idea? If it had been workable, it should have been developed a long time ago! Or, maybe not. Perhaps it is really original. Perhaps testers think in conventional ways and they have therefore not pursued this unorthodox topic. Perhaps it is too alien to the test community. Perhaps there is a general impression that testees usually all are highly motivated and that there is so little variation among them that the motivation topic is not worth pursuing. Be that as it may, the most likely answer is that the idea is, for some reason, unworkable, because otherwise it would have been developed a long time ago. It is not *that* original, after all. Or is it?

Example 2: Emotional intelligence

The concept of emotional intelligence (EI) was introduced by Salovey, DiPaolo and Mayer in 1990 (Mayer, DiPaolo, & Salovey, 1990; Salovey & Mayer, 1990). It was an enormously powerful idea. They found that performance measures of such abilities as identifying emotions could be constructed. Goleman caught the idea blowing in the wind, and wrote a very successful book[7] (Goleman, 1995) where he put forward claims as to the very large importance of emotional intelligence, seemingly credible to many of his readers. Others were eager to jump on the band-wagon and suddenly we had a host of self-report tests of the conventional kind, claiming to be tests of emotional intelligence. Emotional intelligence sounded, in the ears of many, as just the right thing needed in the workplace. The older history of failures in psychology to measure social intelligence had been forgotten (Kihlstrom & Cantor, 2000), and emotional and social intelligence were just assumed to be synonyms, more or less.

Meanwhile, the group of psychologists who had originated the concept of EI refined their performance measures. The first version of their test was subjected to criticism (Davies,

---

[7]. Translated to dozens of languages, and sold in millions of copies all over the world.

Stankov, & Roberts, 1998). It did not seem that the performance measures of emotional abilities converged, after all, and there were grave difficulties in defining what are correct answers to many of the items. The most simple and most often used way to define correct answers is by taking the modal response in a group of subjects, but that solution sounds suspicious to many. Is it really true or reasonable that what most people believe is correct? Psychologists are not used to preaching *that* message in their classes. On the contrary, it is a standard trick in psychology text books to exhibit a number of folk wisdoms which all seem credible but are contradictory.

However, it may be possible to define consensus as *the* correct answer. People are experts on how they feel, so the average judgment of mood or emotional state would be both a consensus and an expert criterion, at the same time. Having realized this, I devised a simple design where the subjects first judged their current and habitual mood, using a scheme developed in the 1970's and since then employed with success many times (Sjöberg, Svensson, & Persson, 1979). Then, they were asked to estimate, or guess, the mood, current and habitual, of the other subjects present and taking part in the same test session. Emotion knowledge was scored by taking the absolute differences between guessed mood of others, and actually observed mood of others (arithmetic means). This process was used in three sessions of testing, in 1999, 2000 and 2001 (Sjöberg, 2001a, 2001b). The emotion knowledge scores obtained in this way correlated to some extent with other measures of emotional intelligence.

In 2002, we analyzed a new set of data and it then dawned upon me that there were three more ways of scoring the mood data. Possibly, these new ideas can be classified as products of serendipity. First, taking absolute differences between own mood and mean mood of the group constituted the basis of measuring deviant emotional reactions. Several analyses revealed this to be quite an interesting new aspect of emotional life, and actually a stronger correlate of such variables as performance EI or social adjustment, than the original emotion knowledge scores.

There were two more ways of scoring the data. First, guessed mood of others could be related to the mean guesses given by the people in the group. This would be a score of how idiosyncratic, or normal, the guesses given by any one person were. In other words, did they think about moods and mood changes in the same way as others did? Second, did their current mood coincide with what people guessed it would be? This could be scored by comparing current mood and the mean guessed mood. We can think of this score as predictability. If the two are close, people have reacted the way others thought they would react. Table 1 summarizes the four types of scoring.

Table 1. The four different possibilities of scoring actual and guessed mood.			
		Target of comparison	
		Mean actual mood	Mean guessed mood
Type of mood rating	Actual mood	1. Normality of emotional reactions	2. Predictability of emotional reactions
	Guessed mood of others	3. Emotion knowledge	4. Normality of conceptions about emotional reactions

Summing up, it was scoring variation number 3 that started our interest in these data. We found much later that variation number 1 seemed even more fruitful. Variations 2 and 4 may serve as starting points for further research and the construction of new scales. They are quite different, in fact. One measures if a person has notions about emotional reactions which are typical for a group, or if he wanders off in a jungle of idiosyncratic thoughts about emotions. The other measures if a person reacts emotionally in predictable ways.

It should be noted that it took three years for me to come to think of all the four scoring possibilities of the mood data. They were initially only devised to make it possible to score emotion knowledge. Will these new ideas lead to some interesting new research? Both variations 2 and 4 could be interesting as starting points for thesis work, and they would perhaps provide new angles on emotional intelligence.

Example 3: Checking it

Here is an approach that is *almost certain* to give good results. Maybe it is not sufficient for a thesis but it could in many cases lend itself to a good start on one.

Psychology is full of half-truths. The student who is starting up a research career has usually read mostly text-books and such books usually do not go into details behind their assertions. They may refer to some classical papers and perhaps more current reviews, but they do not go into the details of research which allegedly support their notions. It is often very rewarding to check these details. Do the papers cited in reviews really say what the reviewer says they do, and how strong is the evidence in fact? It will be a surprise to the student to discover that prominent researchers may cite strange evidence in support of their points, such as other papers which present no new empirical results but merely repeat what others have said already, or papers which do not build on empirical data at all but on simulations. Simulations may have an interest, but hardly as support for an empirical thesis. In addition, there is the old problem of significance versus size of effects, see Sjöberg (Sjöberg, 2003a) for an example and a detailed discussion. Students should be wary of the many cases when empirical findings are based on very small effects, small but significant . What should one do, really, with effects of a few percent explained variance? More and more researchers realize that significance testing is only a first step in establishing a finding, it must also reveal a large effect. Small effects could be due to artefacts, and so could large effects but in those cases the risk is smaller.

Going back to the evidence could carry a thesis a long way. A recent example is described in a paper by Elisabeth Loftus (Loftus & Guyer, 2002a, 2002b). She was interested in the question if sexual abuse in childhood could give rise to repressed memories which would later, in adulthood, be brought back to consciousness. A paper by Corwin and Olafson (Corwin & Olafson, 1997) seemed to demonstrate such a phenomenon in a very forceful and reliable manner. They had interviewed a girl when she was six and had allegedly been subjected to sexual abuse by her mother, and who later had forgotten all about it when she was interviewed once more, ten years later. When shown the original taped interview she came to remember the abuse. Corwin had then shown the tapes to a number of prominent researchers, and they had written papers testifying to their conviction that the case was one of real, documented repression. Only one of the researchers was skeptical. This was a sensation because such repression had previously not been that clearly documented, in spite of its being the basis of many court cases and the subject of long, intense controversy in forensic psychology (Holmes, 1990; Holmes, 1994), see Sjöberg (Sjöberg, 2000-01) for a dramatic Norwegian murder case where a person was convicted on the basis of the alleged return of repressed memory.

Loftus got the idea to track down the people involved in the case, such as the mother of the girl, several doctors and so on. When she did that, she found that the case description was misleading in the sense that many important details had been either left out or had been described in a quite misleading fashion. She also found that it was highly dubious if the girl had really forgotten the events in the time interval between the two sets of interviews. She was said to have talked about the events in the interval. How could she then have forgotten them?

It is not the intention here to go further into this case or the memory wars debate, although it is a goldmine of questionable science (Crews, 1995). I simply want to point out that it may be a good idea to do detailed study of the evidence for assertions made by researchers. Researchers are human beings and as such they are rarely unbiased. They want to support a thesis of one kind or another and they often present biased cases where important details are left out or depicted in a misleading manner. Journals may publish such work because editors and reviewers are human beings, too, and have limited resources available to check in detail all the assertions made in a manuscript. For example, a reviewer of a manuscript rarely will check original sources cited in a paper to find out if they really say what the author of the manuscript asserts. The reviewer may react, but only if he or she has already read those sources, not otherwise. The reviewer will also very seldom have the chance to check the empirical work carried out. Errors in calculations may occur more often than expected. To-day the arithmetic work is done by computers, but errors in entering data and specifying details of the statistical analysis are not unheard of. Such errors can be the cause of what appears to be interesting, unusual or remarkable results.

Example 4: Reading fiction to get ideas

Fiction is a truly enormous field of creative production and many sharp psychological insights. There is a whole academic field devoted to the analysis of fiction, and we shall not go into it. Instead, my suggestion is to look for ideas in good, or even bad, books. The result may be quite interesting.

An example is afforded by a novel which has become a cult book, *viz*. Rhinehart s *Dice Man* (Rhinehart, 1971). The idea is as simple as it is intriguing. A person makes a pact with himself that he will first specify his options, perhaps six of them, then roll a die, and then he is committed to carry out the act which the die decides for him. What would be the result of such die living ? Would it be beneficial, or would it on the contrary lead to disaster? How about some research on those who have tried it, or are willing to do so? The story is about breaking through various inhibitions.

Here is another idea, which occurred to me when I testified in the Norwegian murder case mentioned above. The suspect had been enticed by the interrogating police officer to write a short story describing the murder and the feelings accompanying it. The suspect had no memory of having committed the murder but he complied with the suggestion. He was, as it turned out, a person with some literary talent. He wrote a short story which was quite convincing, both to the police and later to the court, and even to himself. He started to remember ! He even formally confessed to having committed the very brutal act involved (and later retracted the confession, but it was still believed by the courts). My idea was the following. This young Norwegian boy had literary talent. A good author can write as if he has actually experienced a set of events, as if he is guilty of a crime for example. Take Dostoevsky. His *Crime and punishment* is a masterpiece, and surely gives the impression that the author has lived through these or similar events. Are people who read such stories convinced that fantasy is not enough to produce them? How does the impression arise that a piece of fiction is reality? The same question can be asked about the theater or movies. Hollywood has today enormous skill in making all sorts of impossible events seem very real. Is the illusion in fact a destructive force, making people believe in all sorts of superstitions and losing contact with reality?

Example 5: Historical case studies

In our risk research program we happened to learn about the first major railway accident in Sweden[8] (Sjöberg & af Wåhlberg, 1996). One of the many interesting insights afforded by that work was the lack of demands to phase out the railways. After all, about ten people had been crushed to death and many more injured. Should such a dangerous technology really be retained? Look at what happened in Sweden in 1979, after the TMI nuclear accident. People were quite determined to phase out nuclear power. (They didn t but that is another story).

It then struck me that a major factor might be if a technology is seen as substitutable or not. Nuclear power can be substituted with other technologies for producing electricity. Railways constituted a quantum leap in transportation technology in 1864. There was no comparable

---

[8]. The background has some interest. We were invited to contribute a review paper on risk perception with reference to the railways who were just then starting up a new high-speed train service in Sweden. In the course of working on that paper (Drottz-Sjöberg & Sjöberg, 1991) we came across the Sandsjö accident in 1864.

alternative in terms of speed, comfort and cost. Railways opened up the world[9].

To test the notion of substitutability, I studied attitudes to technology in three dimensions: global attitude (good-bad), whether too much attention was devoted to the risks of a technology, or too little, and whether the technology should be phased out or its use allowed to expand. These attitudes were related to explanatory variables: voluntariness, novelty, substitutability, risk and benefit. The results for 4 technologies are given in Table 2. The data were collected from a quasi-representative sample of the Swedish population (Sjöberg, Hansson, Boholm, Peterson, & Fromm, 2002), N=294.

Table 2. Regression analysis results (standardized regression coefficient) for four technologies. Largest value in each row in bold face. Stars denote significance levels.					
	Voluntari-ness	Protection possibility	Substitute-ability	Benefit	Risk	R^2_{adj}
- Genetically modified food -						
Global attitude	-0.033	-0.037	0.152**	-0.260***	**0.478*****	0.489
Too much vs too little attention	0.069	-0.073	0.044	-0.160**	**0.503*****	0.404
Phase out vs. expand	-0.086	0.065	0.147*	-0.079	**0.197*****	0.092
- Nuclear power -						
Global attitude	0.109**	0.063	0.246***	-0.291***	**0.378*****	0.647
Too much vs too little attention	0.027	-0.018	0.037	-0.198**	**0.449*****	0.374
Phase out vs. expand	0.137**	-0.091	**0.421*****	-0.074	0.227***	0.522
- Cellular telephones -						
Global attitude	-0.051	-0.002	0.117*	**-0.421*****	0.269***	0.373

---

[9]. There is a common misconception that new technology is perceived as risky just because it is new. We find no such reactions in current research, and in the case study of the Sandsjö accident we found no negative attitudes against railways *per se*, only criticism of conditions which were partly responsible for the accident.

Table 2. Regression analysis results (standardized regression coefficient) for four technologies. Largest value in each row in bold face. Stars denote significance levels.

	Voluntari-ness	Protection possibility	Substitute-ability	Benefit	Risk	R^2_{adj}
Too much vs too little attention	0.078	-0.149*	0.115	-0.042	**0.318*****	0.168
Phase out vs. expand	0.016	-0.022	**0.265*****	-0.224***	0.190**	0.249
- Pesticides -						
Global attitude	-0.020	0.029	0.261***	-0.171***	**0.500*****	0.553
Too much vs too little attention	0.226***	-0.011	0.209***	-0.115	**0.252*****	0.324
Phase out vs. expand	0.045	-0.040	**0.374*****	-0.086	0.217***	0.324

It seems that the idea is workable. Substitutability works the way we expected, and contributes in an important manner to explaining variation in attitudes toward the technologies[10]. Is this type of research of any importance? I think it is. The social and political problems of risk management are enormous and they are partly psychological. Understanding attitude dynamics in cases like this provides a platform for management. Let s look more closely at another historical example.

A special case: a good bad idea

My prime example here is from risk research, and the seminal work by Starr (Starr, 1969). Starr was the first to analyse risk acceptability of technologies in relation to benefits. He found, as might have been expected, a strong relationship at the aggregate level. The more benefits, the higher the accepted (or actually existing) risk levels. Yet, technologies seemed to form two groups which he termed voluntary and involuntary risks. Examples would be driving your own car (voluntary) or riding in a bus (involuntary). The level of accepted risk was about ten times larger in the former case.

These were interesting findings which came to be the basis of a whole new field of social science risk research. Several researchers jumped on the band wagon. It was especially

---

[10]. There were 14 more technologies in the study, and the pattern of Table 2 was supported generally.

attractive to do so because Starr was clearly wrong in one important sense. Voluntariness was not really a good way of interpreting the findings. What is involuntary about riding a bus? Several alternatives seemed to fit better to our intuitions, such as control . Be that as it may, and the debate is still going on (Sjöberg et al., 2002), the important point is that Starr had hit upon a gold mine of interesting and important problems, and that it was obvious that things needed to be improved upon. Therefore, his achievement was very fruitful. It was a good bad idea . There are many other examples. The crucial point is that a fruitful new field is defined, and that other researchers can see how they can contribute.

More examples of not-so-good ideas

In the present section, I develop a few examples of ideas that are unlikely to lead to successful research. Three different types of bad ideas are identified.

Example 1: Personality

It is quite common that students want to study a phenomenon in relation to personality . The reason is probably at least partly the fundamental attribution error , i. e. the tendency to interpret the behavior of others in terms of their enduring traits and dispositions (Jones & Nisbett, 1970-71). The mistake here is that personality has never been found to account for more than a minor fraction of behavior, in spite of all the convictions to the contrary, and common sense (Mischel, 1968). People vary from one situation to the other. These weaknesses are true of all work on personality, including trait and typology approaches, but typologies are worse than trait approaches. It is very tempting to think in terms of typologies, and much has been written about various approaches, starting in Antiquity and the famous four temperament types of Empedocles and Hippocrates (Adams, 1972). This way of thinking is the basis of some current lucrative commercial applications (Sjöberg, 2000b). The problem with typologies is that people differ, yes, but not in such a simple manner that they can be classified into a small number of types. It is always found that only a few extreme people fit the typological scheme, all others are mixed types .

Example 2: A criterion defined scale

In our work on EI we have found only modest relationships between performance based measures and self-report scales. These findings are similar to the experience of several other researchers. They are problematic because they throw some doubts on the validity of the concept, as it is currently operationalized. I did, however, have very extensive data and they could possibly be used to discover any systematic relationships between the two sets of concepts, or so I thought.

Over 1000 self-report items were available in two data sets, together with a performance measure based on judgments of emotions in vignettes describing social problem situations. The latter were scored according to a consensual scoring key, i. e. the most common response was scored as correct . I then correlated each of the 1000+ self-report items with the performance score in one sample, and found 85 items which correlated significantly. The sample was of modest size, N=41, so fairly high correlations were required to pass as

17

significant . These 85 items were subjected to item analysis and had an excellent internal consistency reliability, above 0.9 in terms of Cronbach s alpha. The total score based on the 85 items correlated 0.7 with performance based EI! Looking at the individual items they were heterogenous but I thought I could see some psychological commonalities in them...

Before establishing the scale , however, a necessary check was a cross validation on a new sample. I had such a sample, N=190. The scale turned out to have a lower alpha in the new sample, 0.7, but still fairly respectable. However, when correlating the scale with EI performance, validity dropped to 0.1! Hence, it was all just a way of capitalizing on chance. The 85 items which had respectable correlations with EI perfomance in the first sample did so just because of random errors, apparently. When many, many tests of significance are made, some are bound to come out as significant , just due to random errors. The test using cross validation is very simple and easy to understand, and quite necessary. A shrinkage is always to be expected, although I was surprised to see how important it was in the present case.

It should be added that even if the scale had held up under cross validation, it would have been of doubtful value. Items selected in this way tend to be heterogenous and make little psychological sense. They may be useful for prediction but lead to no improved understanding.

Example 3: Job satisfaction

People who like their jobs perform better and more, yes? The answer is no, they don t. There are literally hundreds of studies which show that people who like their jobs *don t* perform more, or better (Iaffaldano & Muchinsky, 1985). Once this is pointed out, it seems obvious, like so much else. I may like my job just because it does not demand too much from me. Performance is a different thing, related to the will to work (Björklund, 2001; Sjöberg & Lind, 1994), which is entirely different from job satisfaction.

The bad idea involved here is repeating the basic study over and over again. It seems so attractive that people who like their jobs should also perform better, and Hackman and Oldham suggested theory of job motivation (Hackman & Oldham, 1976) which was based on this idea. According to them, certain job characteristics create job satisfaction (true) and also higher performance (false). There was a big political investment in this line of thinking, which may explain why so many studies were carried out, in spite of repeated failures.

How good ideas differ from not-so-good ideas

The not-so-good ideas have some properties in common. First, they are not developed on the basis of current research, but on common-sense notions.

Second, they are not motivated by true research interest but by something else, such as interest in psychotherapy. There is nothing wrong with that kind of interest, it just is not research interest.

Third, they contain little in the way of intellectual challenge, and thereby also little promise

for further development, or arousing the interest of other researchers.

Fourth, they involve the repetition of old ideas that have already been tried and found more or less useless.

Negating these properties by no means guarantees that one has a good idea. For example, there may be intellectual challenge in an idea, still no really interesting empirical applications. It is only by trying out ideas in practice that we can decide if they are really good.

In my experience, such trials very often bring in new and unexpected aspects, which then, in turn, may be useful for further research. The research process should take place in interaction with the empirical data and new ideas are, in the ideal case, formed on the basis of unexpected results. This is very natural and in a way simple. One just has to have an open mind and look for the unexpected.

The received view of the research process is that it is or should be a question of *hypothesis testing*. Hypotheses should be derived from theory, hence research should be about testing theories. What is wrong with that?

The mistake is the implicit assumption that there are many very interesting theories and that all research should be geared towards testing them. Psychology does not yet have many interesting theories. We need ideas about phenomena and empirical principles, and in a later phase there will be theories. Research needs to be explorative and hypotheses must be very tentative.

How do I get good ideas?

This is a question about creative problem finding and problem solving (Runco & Chand, 1995). Much has been written about the topic, and it is at the present not fully understood. Guilford is famous for having awoken interest in creativity in a presidential address to the American Psychological Association in 1950 (Guilford, 1950). Before that time, there was very little psychological research on the topic, maybe partly because the field was dominated by behaviorism and animal models. Rats are not creative - the topic just does not arise. It is now 50 years since Guilford s address and there is more work on creativity, but it is still a relatively neglected area (Sternberg & Lubart, 1999).

One of the best ideas that Guilford had was that of divergent thinking (Guilford, 1975). A major point is the notion that criticism should be suspended in the phase of generating many ideas, most of which are later to be deleted. The few really good and original ideas are kept. People vary greatly in how good they are at producing many ideas. Some tricks may help, though. A simple approach is to use random input, such as random words or pictures. Most of these methods are probably not well suited for generating research ideas, however. Let s take a closer look at that particular problem.

Simple, everyday problems call for quick and dirty fixes. What would be a good Christmas gift for a person I know? Or what should we serve for dinner to people we have invited? Or what should I wear for that party I am invited to? Answers to such questions require creative

handling of a mixture of general cultural knowledge and idiosyncratic knowledge about particular people, places and situations. Research is different, in different ways for different disciplines. Psychology certainly is related to general cultural knowledge, and it may be related to knowledge about particular people known to the researchers, or maybe people he or she has read about, or seen in movies. But mainly, the substrate to develop is found in the research literature. The way to get ideas is therefore to read relevant previous work, and to do so with an open and critical mind. Given that the researcher has this pertinent and relevant knowledge he or she can start to generate ideas.

Although trait approaches have had only very limited success generally, it should be mentioned that there are some personality factors of importance for creativity. An important factor in creativity is intellectual openness, one of the factors in the five-factor model of personality (McCrae & Costa, 1987). A negative factor is perfectionism and compulsive orderliness. Creative people are relatively low in two other factors: conscientiousness and sociability. They develop in their own directions of work and often do so by themselves, not seeking the cooperation of others. They tend to be introverted at times. They also tend to be emotionally unstable. All these negative characteristics must not be too pronounced, of course, because they then will be obstacles to creative performance.

Good ideas may be found in many places, in case you don t produce them yourself. There is a saying: If you want to get a new idea, read an old book . This is absolutely right. A discipline such as psychology has a short memory. Maybe this is so because many researchers entertain an illusion of complete and perfect cumulation of knowledge in the field. Current versions of anything are then necessarily compilations of all that is good in previous work, adding some new notions. But reading old books also gives a valuable perspective on just how slow progress in our field is. An example: the first book in Swedish on witness psychology was published in 1933 by Arvid Wachtmeister (Wachtmeister, 1933). It is an excellent piece of work, in good contact with German and French work from the first decades of the 20th century and with many interesting case studies and shrewd insights into the intricate problems of human testimony in court. Yet, it seems to be completely forgotten to-day, and so is most of the German and French extensive research on witness psychology as well. Surely, anyone who wants to do research in forensic psychology could get new ideas from this text and the work that it cites.

Managing creative research

Consider now how creative research should be managed. Researchers are oriented towards their own world of ideas. These ideas may or may not have practical applications. Important practical applications require innovative research. The problem in managing creativity is that managers want quick fixes, and are suspicious of researchers. They suspect that researchers have a good time and enjoy themselves, and get paid for it. This is somehow wrong. Work should be a burden, as apparently was the belief of people of my school who conducted a survey about graduate students attitudes. They never even asked whether the students enjoyed being creative.

We find these counter-creative forces at work everywhere in applied research. The great successes of such work illustrate the exact opposite. Swedish pharmaceutics company Hässle

was enormously successful under the leadership of Ivan Östholm (Östholm, 1995). Östholm was not a researcher himself, but he was sensitive to the needs of the academic researchers he commissioned to work. Their most famous achievement was Losec, the world s most successful drug and an enormous commercial success story. What were the needs of the researchers? It was to be sponsored for doing basic research and to publish in reputable journals. This was the way for them to make a career, not with Hässle but in the academic community. The arrangement worked enormously well, but it seems that few have been influenced by this lesson. I suppose the message is too alien to most managers. Swedish Government authorities do sponsor social science research, but they demand compliance to fuzzy programs which seem to be geared towards discouraging basic research. This is counterproductive. They will not get the best researchers to apply, and they will not get cutting-edge research results.

The Losec case illustrates the importance of *intrinsic motivation*. A large amount of research has shown that extrinsic orientation towards rewards such as money will be detrimental to creativity (Amabile, 2000; Amabile, Collins, Conti, & Phillips, 1996). Part of the problem may be that managers do not require the right kind of achievement. For example, if a university or college does not reward its professors for publishing internationally, many of them will not do so. It is hard and often frustrating to try to compete in the international arena. Why do it if not required? This attitude is even worse for the graduate student who will not learn in time the tricks of the trade. He or she will be pursuing what can be called a local career, if that strategy is encouraged. As noted above, the result is that meaning is lost. Research must contribute to the international community or it is a waste of time and money.

The main force creating good ideas is intellectual excitement and emotional support from the environment. It takes time, of course, to get the really good ideas and all of them will not hold up under empirical testing. However, a highly critical and competitive attitude in the environment will kill many ideas even before they are hatched. The very process of deepening the ideas, and shaping them to fit realities is interrupted or not even started. Many academic environments are like this. Doctoral students may begin to look upon each other as competitors (a very tempting and natural thing to do), and if their thesis advisers are pursuing the Holy Grail of Perfection, very little will come out of many years of work. The climate in a research team needs to be accepting and strongly encouraging, and criticism must be dealt with in a very delicate manner.

It is possible that a multicultural environment is fruitful (Ducker & Tori, 2001). However, there is no guarantee for that in psychology, because most research in our discipline is culturally mute. It is also possible that gender plays a role in psychological research, but I don t believe much in that possibility. Women tend perhaps to choose different fields of study than men do, but I see no tendency for them to think and do research in different ways. There is no female way of knowing which is different from the male one, not in scientific research. (There may be in everyday social and emotional knowledge, but that is another story).

Success or failure?

Why opt for a research career? I assume that the goal is to be successful in research, which

means many things. It means getting and developing good ideas, which turn out to work empirically. Do good ideas, and original ideas, lead to fame and international recognition? By no means. There is a lot more required.

The main route to recognition is international publications in the leading journals in the specialty. It is easy to locate those journals, not easy to have papers accepted by them. Let me first say a few words about the orientation of the journals.

There are two groups: general psychology journals and journals specialized according to interdisciplinary contents. Risk research, to take that example once more, can be published in general psychology journals such the *European Psychologist* or the *Journal of Applied Social Psychology*. I call these general journals because they specify their orientation in terms of scientific discipline, not contents. Then there are journals which are interdisciplinary and specialize in risk, such as *Risk Analysis* or the *Journal of Risk Research*. These journals are about risk, but many disciplines are represented, and not only social or behavioral science, also natural science, engineering and medicine. The people who read them do so because they have a professional investment in risk matters. Psychologists who read psychology journals are rarely interested in risk questions. Hence, to get interested readers, the journals specializing in the substance of research are to be clearly preferred. The exception is the researcher who is strongly bent on theoretical and basic scientific work, here it may be appropriate to try for the best psychology journals. My own experience confirms strongly that the specialized journals are to be preferred in other cases, if you want to get interested readers.

All this talk about journals is essential for the young researcher because funding of research projects, and appointments to desirable jobs in universities, are based more and more on journal publication achievements. This is the way it should be, because one must break the local in-group dynamics, in which local reports or books are all that is required. There are still environments like that in Scandinavian universities but they fight a losing battle. Psychology in Sweden turned decidedly international in its orientation in the 1960's and a recent evaluation of psychology and other disciplines in major research nations showed it to do quite well (May, 1997).

There are excellent handbooks about writing manuscripts for scientific journals, most notably the APA publication manual (American Psychological Association, 2001). They tell you to write clear, short papers which stand a chance of being accepted. The chance is still small, of course. But even if it is only 5 percent, or less, it is still quite possible to succeed, provided many attempts are made (Sjöberg, 2003b). The student must develop a thick skin and not be discouraged by many failures. Everybody fails, including experienced professors. Also, the letters from editors should be carefully read. A paper is almost never accepted without revision. What looks like failure may simply mean that the editor wants a revision and then is ready to accept the paper.

When a journal is selected, it is a good idea to look closely at the kind of papers they have recently been publishing lately. You may even find some interesting pieces of work that you had missed. You will get a feeling for the flair of the journal, the style they use, and the kind of topics they deal with. If your own list of references already lists a number of papers from that journal, you have probably made the right choice.

Discussion and conclusions

Several topics of this paper have been rather seldom discussed. This is a bit strange, because judgments are made all the time on how good or bad ideas are, e.g. by selection committees for psychology chairs or by committees working for research councils in assessing applications for funding. There seems to be very little reflection in these circles on just what it is that they are judging. There may indeed be great variation in judgments. Also, there seems to be little reflection and discussion about how to create and sustain a creative social climate in a research group.

The present paper succeeds if it manages to stimulate some of its readers towards fruitful reflection on their research, and on ways to renew their approaches. That remains to be seen, of course. The ideas of the paper may be correct, still it may fail to induce the right mind set for fruitful research. Or the ideas may be simply wrong. Perhaps very different criteria could be applied to decide if an idea is good or bad, and perhaps the present notions will not be helpful in generating good ideas.

It is obvious that ideas cannot be commanded. They come or they do not come. I have tried to show in this paper that getting ideas is a matter of openness. This is largely a question of attitude and hence not too hard to change. It all boils down to taking in as much information as possible, and forming a network of inter-related associations. The process takes time, and cannot be done as an ordinary job. It is a question of the stream of interests , of relating and integrating various parts of life into one or several themes which should then start to interact, strengthen each other and give rise to new themes. It is not really hard, but it takes time and commitment, and openness.

It was noted in the introduction that only 20 percent of the graduate students in social sciences in Sweden ever manage to finish with a Ph. D. This is true for psychology as well. What are the reasons? A reform of graduate study back in 1969 was based on the complaint that those who finished with a Ph. D. degree were as old as 40 years. The Government set about to do something about that and a very extensive program of reform was launched. Thirty-three years later the age of finishing students was even a few years more than it was in 1969. Almost no assessment was made of the reform until 1996, and virtually nobody discussed what could be wrong with it. In 1996 the National Audit Office finally did assess the reform and found it to have utterly failed [Riksrevisionsverket (Swedish National Audit Office), 1996]. Average time to complete graduate education was about 10 years in the 1960's, and so it was in the 1990's. The question is why the reform failed.

The National Audit Office stressed primarily the economic side and suggested that more

students should get the chance to pursue full-time[11] graduate work[12]. The recommendation was presumably based on the fact that students themselves argued that the major obstacle to completing graduate work was inadequate economic support. However, there was no analysis of whether completion or time to completion was related to financial support. Having seen several failures in spite of generous economic support, I was skeptical about this strong emphasis on economics. Maybe it is a contributing factor, but there must surely be several others.

In 1998 a reform of graduate education was implemented. It was now required that all new graduate students should have a guarantee of full economic support for 4 years and that universities should carefully monitor their progress and see to it that they received adequate supervision. Since the total resources were not radically increased, this meant that much fewer students could be admitted and that, in turn, made it hard to carry through the usual program of doctoral courses. Was this reform - surely well-meaning - really what was needed? How important are economic resources[13]? Nobody seems to know.

However, it is hardly likely that economic factors are sufficient to alone bring about the desired increased completion rates and decreased study times. Let me suggest a few other themes which may be of relevance.

1. If 80% drop out, 20% actually get their Ph.D. How do they manage? Is it because they had especially competent thesis advisers, or that the process of research guidance involved very close monitoring of their work and integration into research teams? Or were they people with particularly high talent, motivation and creativity? If the latter is true, it is possible that improvements require new selection procedures rather than new supervision principles or, possibly, both. However, how should one go about selecting students? Extensive experience with psychological tests is not encouraging (Sternberg & Williams, 1997).

2. In some other countries, perhaps especially the USA, things seem to function much better than in Sweden. Or is this an illusion, something merely true of the leading US universities? It would be interesting to have some discussion about this, and to learn whether US principles of graduate education are in fact better. Swedish experience is limited when it comes to the establishment of elite institutions, in spite of resources being available, in principle, for doing so (Sjöberg, 1999c).

---

[11]. Conditions varied, and many of the students receiving these scholarships also had to fulfill some teaching and administration duties. There seem to be no data available on how common this was, and how large a fraction of full time was involved.

[12]. Another suggestion concerned continuity of supervision, which they found to be in conflict with supervisors own need for time free from all teaching duties. I concentrate here on the major factor, the economic one.

[13]. At the time of writing (May 2003) another follow-up has been commissioned and it will presumably report on the results of the 1998 reform. A report is due in the end of 2003.

3. The article does not discuss the role of the public critical discussion of the thesis, in Swedish "disputation" or Public Defense. This is a traditional custom still upheld in Sweden. Even if almost nobody fails that test, it is a considerable emotional burden for most students. Maybe some less dramatic procedure could speed up the process and also increase output from graduate education. I am thinking of the UK system. In addition, the printed version of the thesis would be finalized *after* the final scrutiny, not before it, and the resulting thesis would be improved.

4. The long delay and the enormous drop-out may be caused by thesis work, or course work, or both. It is possible that some parts of course work (which usually requires about half of the stipulated time) should be devoted to courses directly aimed at supporting the students' research, such as courses about publishing articles in international journals. I am aware that some courses already have aims in this general direction, but more could be done.

5. Delay and drop-out could also be related to the job market for graduates. Some students may give up simply because they do not believe they could get a good and well-paid job with a Ph. D. while they see alternatives with a BA or an MA Others stay on as graduate students more or less on purpose and wait for one of the very few vacant openings in their department. Very few want to or can succeed in switching to another department since most departments give priority to their own people, a very unsound system. Compare here with the US system. If the Government *really* wants more Ph. Ds why don't they create jobs for them?

6. There are implicit, sometimes explicit, norms about theses which are contrary to the official norms. People have not really changed their ideas about thesis work very much since 1969. A thesis which is done in the stipulated time tends to be scorned, and the thesis adviser loses face. The adviser who is uncertain about his or her own status and job prospects has little to lose by increasing the demands on his on her students.

7. Finally, there is ambivalence in the system about ideas and originality. The Government wants to downplay those aspects and sees graduate education mainly as acquiring certain skills of applying scientific methods. Somehow, the creative side of research is assumed not to be important or to take care of itself. Yet, it does exist and the informal norms in many contexts require creativity from the student. Not least does the student him- or herself require creativity. Failure and delay are probably often due to the inept handling of this aspect of graduate work. The student who has his or her very own idea is the one burning with enthusiasm, for that student research is no burden and the degree will be finished.

On the other hand, the supervisor may suggest and formulate both the basic idea and then one idea after the other in the course of work. This may be helpful and even necessary. The risk is that the student does not learn to develop creativity, and may even be deterred from doing so, especially if the supervisor is obviously good at generating interesting ideas. The supervisor also has career aspirations. He or she may insist on being a co-author, even first author, on some and all of the papers making up the thesis. This is natural and common practice in many disciplines, more or less unheard of in others. In the former case, motivational problems may arise. In the latter case, there may be a relational problem because the supervisor may resent that the student poses as the originator of ideas that really were the supervisor s, not the student s. To make things even more interesting, it is frequently hard to decide who was the

originator of an idea, because several people can have it at the same time, or memory fails them. Be that as it may, I believe it is good practice to be explicit about these matters, and bring them up as early as possible.

The main message of the present article is that (a) all good research, including thesis work, requires creativity, and (b) creativity can be taught and stimulated in various ways, some of which are well known to be effective. This may sound like a trite conclusion, but in the light of the practice implemented in our university system, and especially in applied research, it is not. The National Audit Office did a good job in pointing to the problems, but their preferred solutions were typical for bureaucrats and politicians: very concrete, and probably missing the most important points having to do with basic psychological processes. Administrators often seem to favor a view of research and research education as learning and applying certain tools . This, too, is necessary, of course - but if you lack ideas about good problems to apply the tools on, then what is the meaning of the whole enterprise?

References

Adams, F. (1972). *Genuine works of Hippocrates*. New York: Krieger.

Amabile, T. M. (2000). How to kill creativity. *Harvard Business Review*(September-October), 77-87.

Amabile, T. M., Collins, M. A., Conti, R., & Phillips, E. (1996). *Creativity in context. Update to the social psychology of creativity.* Boulder, CO: Westview Press.

American Psychological Association. (2001). *Publication manual of the American Psychological Association, 5th Ed.* Washington, DC: American Psychological Association.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*(1).

Björklund, C. (2001). *Work motivation - Studies of its determinants and outcomes.* Stockholm: EFI.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 3*, 186-190.

Corwin, D. L., & Olafson, E. (1997). Videotaped discovery of a reportedly unrecallable memory of child sexual abuse: Comparison with a childhood interview videotaped 11 years before. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children, 2*, 91-112.

Crews, F. (1995). *The memory wars. Freud's legacy in dispute*. New York: The New York Review of Books.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology, 24*, 349-354.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989-1015.

Drottz-Sjöberg, B.-M., & Sjöberg, L. (1991). *High speed trains and the perception of risk* (Rhizikon: Risk Research Report 2): Center for Risk Research, Stockholm School of Economics.

Ducker, D. G., & Tori, C. D. (2001). The reliability and validity of a multicultural assessment instrument developed for a graduate program in psychology. *Professional Psychology: Research & Practice, 32*, 425-432.

Ekman, G. (1958). Two generalized ratio scaling methods. *Journal of Psychology, 45*, 287-295.

Ekman, G., & Sjöberg, L. (1965). Scaling. *Annual Review of Psychology, 16*, 451-474.

Ericsson, K. A. (1999). Creative expertise as superior reproducible performance: Innovative and flexible aspects of expert performance. *Psychological Inquiry, 10*, 329-333.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444-454.

Guilford, J. P. (1975). Relation of divergent-production abilities to verbal and nonverbal IQ's. *Journal of Multivariate Experimental Personality & Clinical Psychology, 1*, 278-284.

Hacking, I. (2000). *The social construction of what?* New York: Harvard University Press.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16*, 250-279.

Holmes, D. S. (1990). The evidence for repression: An examination of 60 years of research. In Singer (Ed.), *Repression and dissociation* (pp. 85-102).

Holmes, D. S. (1994). Is there evidence for repression? Doubtful. *Harvard Mental Health Letter, 10*(12), 2-4.

Iaffaldano, M. T., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin, 97*, 251-273.

Jones, E. E., & Nisbett, R. E. (1970-71). The actor and the observer: divergent perceptions of the cause of behavior. In E. E. Jones & D. Kanouse & H. H. Kelley & R. E. Nisbett & S. Valins & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, NJ: General Learning Press.

Kihlstrom, J. F., & Cantor, N. (2000). Social intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 380-395). Cambridge, UK: Cambridge University Press.

Loftus, E. F., & Guyer, M. J. (2002a). Who abused Jane Doe? Issues, questions, and future directions. *The Skeptical Inquirer, 26*(4), 37-40.

Loftus, E. F., & Guyer, M. J. (2002b). Who abused Jane Doe? The hazards of a single case study. Part 1. *The Skeptical Inquirer, 26*(3), 24-32.

May, R. M. (1997). The Scientific Wealth of Nations. *Science, 275*, 793-796.

Mayer, J. D., DiPaolo, M. T., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment, 54*, 772-781.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28*, 1-14.

McClelland, D. C. (1998). Identifying competencies with behavioral-event interviews. *Psychological Science, 9*, 331-339.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Paris, J. (2000). *Myths of childhood*. New York: Brunner-Routledge.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness and trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197-1208.

Paulhus, D. L., Bruce, N. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100-108.

Paulhus, D. L., & Reid, D. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307-317.

Rhinehart, L. (1971). *The dice man*. London: Talmy, Franklin.

Riksrevisionsverket (Swedish National Audit Office). (1996). *Samhällsvetenskaplig forskarutbildning. "Four years -not for years". [Research education in the social sciences]* (1996:52). Stockholm: RRV (Swedish National Audit Office).

Runco, M. A., & Chand, I. (1995). Problem finding, evaluative thinking, and creativity. In M. A. Runco (Ed.), *Problem finding, problem solving, and creativity* (pp. 40-76). Norwood, NJ: Ablex.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and*

Personality, 9, 185-211.

Sapadin, L. (1999). *Beat procrastination and make the grade. The six styles of procrastination and how students can overcome them.* Harmondsworth, Middlesex: Penguin.

Scharnberg, M. (1993a). *The non-authentic nature of Freud's observations. Vol I. The seduction theory.* Stockholm: Almqvist & Wiksell.

Scharnberg, M. (1993b). *The non-authentic nature of Freud's observations. Vol II. Felix Gattel's early Freudian cases, and the astrological origin of the anal theory.* Stockholm: Almqvist & Wiksell.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.

Sjöberg, L. (1966). A method for sensation scaling based on an analogy between perception and judgment. *Perception and Psychophysics, 1*, 131-136.

Sjöberg, L. (1971). Three models for the analysis of subjective ratios. *Scandinavian Journal of Psychology, 12*, 217-240.

Sjöberg, L. (1976). Internationellt inriktad psykologisk forskning i Norden. *Nordisk Psykologi, 28*, 209-213.

Sjöberg, L. (1977a). *Kritisk psykologi och psykologisk kritik (Critical psychology and psychological criticism).* Stockholm: Psykologiförlaget.

Sjöberg, L. (1977b). Är internationellt inriktad forskning dålig forskning? (Is internationally oriented research bad research?). *Nordisk Psykologi, 29*, 287-288.

Sjöberg, L. (1981). On the homogeneity of psychological processes. *Quality and Quantity, 15*, 17-30.

Sjöberg, L. (1982). Logical versus psychological necessity: A discussion of the role of common sense in psychological theory. *Scandinavian Journal of Psychology, 23*, 65-78.

Sjöberg, L. (1983). Defining stimulus and response: An examination of current procedures. *Quality and Quantity, 17*, 369-386.

Sjöberg, L. (1985). *Study interests* (Swedish Research on Higher Education 1985:2). Stockholm.

Sjöberg, L. (1987). Conceptual and empirical status of mental constructs in the analysis of action. *Quality and Quantity, 16*, 125-137.

Sjöberg, L. (1997). *Studieintresse och studiemotivation.* Stockholm: Svenska Arbetsgivareföreningen och Institutet för individanpassad skola.

Sjöberg, L. (1999a). On appropriate targets for criticism in psychology. *Scandinavian Journal of Psychology, 40*, 85-88.

Sjöberg, L. (1999b). Psykologisk forskningsmetodik och praktik: tre haverier. (Methods and practice of psychological research: three disasters). *VEST: Tidskrift för Vetenskaps- och teknikstudier, 12*(3), 5-25.

Sjöberg, L. (1999c). Så genomförs en konsekvent uppbyggnad av en elithögskola. *Universitetsläraren*(14), 16-17.

Sjöberg, L. (2000a). The methodology of risk perception research. *Quality and Quantity, 34*, 407-418.

Sjöberg, L. (2000b). *Vad säger egentligen psykologiska test? (What do psychological tests really tell you?)* (SSE/EFI Working Paper Series in Business Administration 2000:9). Stockholm: Handelshögskolan, Sektionen för ekonomisk psykologi.

Sjöberg, L. (2000-01). Psykodynamisk psykologi i rättsväsendet: Ett mordfall i Norge.

(Psychodynamic psychology in the courts. A case of homicide in Norway). *Juridisk Tidskrift, 12*(3), 735-754.

Sjöberg, L. (2001a). *Emotional intelligence measured in a highly competitive testing situation* (SSE/EFI Working Paper Series in Business Administration 2001:13). Stockholm: Stockholm School of Economics.

Sjöberg, L. (2001b). Emotional intelligence: A psychometric analysis. *European Psychologist, 6*, 79-95.

Sjöberg, L. (2002). *The distortion of beliefs in the face of uncertainty* (SSE/EFI Working Paper Series in Business Administration 2002:09). Stockholm: Stockholm School of Economics.

Sjöberg, L. (2003a). Distal factors in risk perception. *Journal of Risk Research, 6*, 187-212.

Sjöberg, L. (2003b). The psychology of succeeding with very difficult things. *Unpublished manuscript.*

Sjöberg, L. (in press). Hur internationell är svensk beteendevetenskap? *Universitetsläraren.*

Sjöberg, L., & af Wåhlberg, A. (1996). *Sandsjöolyckan. (The accident at Sandsjö)* (Rhizikon: Rapport från Centrum för Riskforskning 6). Stockholm: Centrum för Riskforskning.

Sjöberg, L., & Dahlstrand, U. (1987). Subject matter attributes and study interests in post-secondary education. *Higher Education, 16*, 357-372.

Sjöberg, L., & Engelberg, E. (2002). Measuring and validating emotional intelligence as performance or self-report. *Unpublished manuscript.*

Sjöberg, L., Hansson, S.-O., Boholm, Å., Peterson, M., & Fromm, J. (2002). Attitudes toward technology and risks. *Unpublished manuscript.*

Sjöberg, L., & Lind, F. (1994). *Arbetsmotivation i en krisekonomi: En studie av prognosfaktorer. (Work motivation in a crisis economy: A study of prognostic variables)* (Studier i ekonomisk psykologi 121): Institutionen för ekonomisk psykologi, Handelshögskolan i Stockholm.

Sjöberg, L., Svensson, E., & Persson, L.-O. (1979). The measurement of mood. *Scandinavian Journal of Psychology, 20*, 1-18.

Smedslund, J. (1991). The pseudoempirical in psychology and the case for psychologic. *Psychological Inquiry, 2*, 325-338.

Sokal, A., & Bricmont, J. (1998). *Fashionable nonsense. Postmodern intellectuals' abuse of science.* New York: Picador.

Starr, C. (1969). Social benefit versus technological risk. *Science, 165*, 1232-1238.

Sternberg, R., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). Cambridge: Cambridge University Press.

Sternberg, R. J., & Williams, W. M. (1997). Does the graduate record examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist, 52*, 630-641.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: Wiley.

Tiner, J. H. (1981). *Isaac Newton: inventor, scientist, and teacher.* New York: Mott Miel.

Wachtmeister, A. (1933). *Vad är sanning? Vittnespsykologins grunddrag.* Stockholm: Natur och Kultur.

Windschuttle, K. (1996). *The killing of history. How a discipline is being murdered by literary critics and social theorists.* Paddington NSW: Macleay.

Östholm, I. (1995). *Drug discovery : a pharmacist's story.* Stockholm: Swedish

Pharmaceutical Society (Apotekarsocieteten).